2022

# STATE OF THE MAP

FIRENZE

# Proceedings of the Academic Track

## Editors:


**Marco Minghini**

**Pengyuan Liu**

**Hao Li**

**A. Yair Grinberger**

**Levente Juhász**


DOI: 10.5281/zenodo.7004791

# Contents

# OpenStreetMap, beyond just Data: The Academic Track at State of the Map 2022

A. Yair Grinberger[1,*], Pengyuan Liu[2], Hao Li[3,4], Levente Juhasz[5] and Marco Minghini[6,†]

[1] Department of Geography, The Hebrew University of Jerusalem, Jerusalem, Israel; yair.grinberger@mail.huji.ac.il
[2] Urban Analytics Lab, Department of Architecture, National University of Singapore, Singapore; pyliu93@nus.edu.sg
[3] Professorship of Big Geospatial Data Management, Technical University of Munich, 85521, Ottobrunn, Germany;
[4] GIScience Chair, Institute of Geography, Heidelberg University, 69120 Heidelberg, Germany; hao.li@uni-heidelberg.de
[5] GIS Center, Florida International University, Miami, FL, USA; ljuhasz@fiu.edu
[6] European Commission, Joint Research Centre (JRC), Ispra, Italy; marco.minghini@ec.europa.eu

* Author to whom correspondence should be addressed.
† The views expressed are purely those of the author and may not in any circumstances be regarded as stating an official position of the European Commission.

OpenStreetMap (OSM) is a project and a community, or community of communities [1], geared towards producing a free, editable, and global geographic database to which anyone can contribute. With more than 8.8 million registered users contributing more than 7.8 billion data points as of 8 August 2022 [2], it has attracted attention across various spheres, from tech giants [1], through governmental organizations and NGOs [3,4], popular media [5], social activists [6], to the academic world [7–9]. The latter is reflected not only in a large corpus of scientific publications relating to OSM, but also in the establishment of the *OSM-science* mailing list [10], dedicated to correspondence on academic studies of OSM, and since 2018 – in the inclusion of a dedicated Academic Track in the annual State of the Map (SotM) conference, the global meeting of the OSM community [11]. The proceedings of the Academic Track at the SotM 2022 conference, taking place in Florence, Italy on August 19-21, 2022 [12], include 19 short papers corresponding to 9 talks and 10 lightning talks presented at the conference. These talks join 49 talks from the previous 4 Academic Track editions as an example of the continued interest of the scientific community in OSM.

The study of OSM is a study of a research object that keeps on evolving and changing [13]. This opens for multiple ways to approach it. However, a classification of recent OSM-related publications [7] shows that some ways are more dominant than others, with the vast majority of papers following data-centric approaches. Unsurprisingly, this theme is also evident in some of the studies included in these proceedings. Several abstracts present domain-specific applications of OSM data, combining OSM data to derive public urban green spaces [14], plan sustainable transport infrastructures [15], map detailed floor plans from digital building models into OSM and back out [16], and using OSM to

Grinberger, A.Y., Liu, P., Li, H., Juhász, L., & Minghini, M. (2022). OpenStreetMap, beyond just Data: The Academic Track at State of the Map 2022.
In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.). Proceedings of the Academic Track at State of the Map 2022, Florence, Italy, 19-21 August 2022. Available at https://zenodo.org/communities/sotm-22
DOI: 10.5281/zenodo.7004424

understand global urban green accessibility [17]. Other studies explore the potential for integrating OSM data with other data sources, e.g. combining OSM water tags with Water Point Data Exchange (WPdx) to improve the mapping of rural water infrastructures in Africa [18], or working towards a knowledge-graph (i.e. Wikidata) integrated OSM dataset [19]. Analyses of data quality are not absent here with Herfort et al. [20] assessing OSM building completeness in urban areas across the world, Camboim et al. [21] investigating the impact of grid cell size for OSM data quality analysis, and Dickinson et al. [22] assessing OSM data quality regarding network navigability in areas where high levels of corporate contributions were observed. A final group of abstracts emphasizes data accessibility, either by supporting mapping activities or data analysis: Anderson & Omidire's [23] development of the Analysis-Ready Daylight OSM Distribution makes global OSM data analysis easier and faster; Vestena et al. [24] introduce a new open-source QGIS Plugin named "OSM SidewalKreator" to help the OSM community to better draw the geometries of sidewalks, crossings, and kerbs in an automatic manner; Schott et al. [25] proposed a workflow to enhance multi-label remote sensing image classification by automatically extracting OSM multi-label training data and verifying them via a feedback-loop in the Tasking Manager projects.

Yet, as noted above, OSM is not just a database - it is the cumulative result of the actions of individuals, organizations, and communities, all being a fundamental part of what OSM is. Hence, OSM is also a social product in which interpersonal, organizational, and behavioral dynamics play a pivotal role [26]. The implication of this is that OSM is a system that extends beyond itself with flows of inputs, people, and resources coming from other systems into it and back out. This social perspective, while existing in the literature, had historically received much less attention [7]. Surveying the abstracts included in these proceedings suggests that the tide may be turning, with 7 of the abstracts detailing research that is socially-oriented in one of several ways. First, 3 abstracts present applications of OSM data that are specifically geared towards social causes, exploring the potential of participatory mapping using a dedicated mobile app based on OSM for promoting geo-literacy among high-school students [27], for analyzing the accessibility of urban spaces for the visually impaired thus prompting equal mobility and walkability in the city [28], or for mapping vulnerable spaces such as refugee camps, using open drone imagery collected as part of the activities of the Humanitarian OpenStreetMap Team (HOT) [29]. Generally, humanitarian efforts within OSM, such as in the last example, seem to induce more socially-oriented research, as seen in Solís' exploration of the way the YouthMappers movement within OSM uses universities as hybrid organizations to navigate between global humanitarian efforts and students' local motivations [30], or Steele's anthropological study of OSM [31] (probably the first since Lin's pioneering work [32]) that uncovers the above mentioned flows through the notion of 'supply chains'. The work of Shrestha et al. [33] joins the one about YouthMappers mentioned above in studying effects on participants by showcasing that mapping skills training of recent high-school graduates and undergraduate students have long-term benefits for youth. Finally, Juhász and Mooney [34] shine a unique light on social dynamics by exploring the meaning for OSM of null island, the fictitious place located at the origin of the WGS1984 coordinate system to which much data is erroneously allocated.

The differentiation made here between data-oriented and social research is not meant to suggest that one is better or more required than the other. This differentiation may not even be real or beneficial given the extent to which contribution processes and data are interlaced [35] and the awareness of many authors to the social side of OSM, sometimes through direct interactions [7]. Hence, the growing attention to the social aspect of OSM is a positive sign showing that the scientific endeavor termed as OSM science [36] is further developing and maturing. But it also shows that it still has room to grow, promoting interdisciplinary collaborations between researchers that can comprehensively consider both the data and the social aspects of OSM. We use this editorial as an open call for researchers to pursue this direction, further enhancing our understanding of OSM.

## References

[1] Anderson, J., Sarkar, D., & Palen, S. (2019). Corporate editors in the evolving landscape of OpenStreetMap. *ISPRS International Journal of Geo-Information, 8*(5), 232.

[2] OSM Contributors (2022). OpenStreetMap Stats. Retrieved from https://planet.openstreetmap.org/statistics/data_stats.html

[3] Haklay, M., Antoniou, V., Basiouka, S., Soden, R., & Mooney, P. (2014). Crowdsourced geographic information use in government. *Report to Global Facility for Disaster Reduction and Recovery* (*World Bank)*, London.

[4] Minghini, M., Kotsev, A., & Lutz, M. (2019). Comparing INSPIRE and OpenStreetMap data: how to make the most out of the two worlds. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLII-4/W14*, 167–174.

[5] Elliott, V. (2022). Ukrainians are scrubbing open-source maps to keep intel from Russia's army. *Rest of the World*. Retrieved from https://restofworld.org/2022/ukrainians-osint-maps-russia

[6] Quinn, S., & Yapa, L. (2016). OpenStreetMap and food security: A case study in the city of Philadelphia. *The Professional Geographer, 68*(2), 271–280.

[7] Grinberger, A. Y., Minghini, M., Yeboah, G., Juhász, L., & Mooney, P. (2022). Bridges and barriers: An exploration of engagements of the research community with the OpenStreetMap community. *ISPRS International Journal of Geo-Information, 11*(1), 54.

[8] Sehra, S.S., Singh, J., & Rai, H.S. (2017). Using Latent Semantic Analysis to Identify Research Trends in OpenStreetMap. *ISPRS International Journal of Geo-Information*, *6*(7), 195.

[9] Yan, Y., Feng, C.-C., Huang, W., Fan, H., Wang, Y.-C., & Zipf, A. (2020). Volunteered Geographic Information Research in the First Decade: A Narrative Review of Selected Journal Articles in GIScience. *International Journal of Geographic Information Science*, 2020, *34*(9), 1765–1791.

[10] OSM Science mailing list: Discuss Anything Science Related in the OpenStreetMap Ecosystem. Retrieved from https://lists.openstreetmap.org/listinfo/science

[11] State of the Map 2018, Academic Track: Call for abstracts. Retrieved from https://2018.stateofthemap.org/academictrack

[12] State of the Map 2022. Retrieved from https://2022.stateofthemap.org

[13] Grinberger, A. Y., Anderson, J., Mooney, P., Ludwig, C., & Minghini, M. (2021). OpenStreetMap as a multifaceted research subject: the Academic Track at State of the Map 2021. In: Minghini, M., Ludwing, C., Anderson, J., Mooney, P., & Grinberger, A.Y. (Eds.) *Proceedings of the Academic Track at the State of the Map 2021 Online Conference, July 9-11 2021*, 1–5.

[14] Rieche, T., & Hecht, R. (2022). Automated derivation of public urban green spaces via activity-related barriers using OpenStreetMap. In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.) *Proceedings of the Academic Track at State of the Map 2022*, Florence, Italy, 19-21 August, 2022.

[15] Timaite, G., Hulse, J., & Lovelace, R. (2022). OSM for sustainable transport planning. In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.) *Proceedings of the Academic Track at State of the Map 2022*, Florence, Italy, 19-21 August, 2022.

[16] Tauscher, H., Krishnakumar, S., & Heigener, D. (2022). Floor plan extraction from digital building models. In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.) *Proceedings of the Academic Track at State of the Map 2022*, Florence, Italy, 19-21 August, 2022.

[17] Battiston, A., & Schifanella, R. (2022). Understanding and modelling accessibility to public green in large urban centers using OpenStreetMap data. In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.) *Proceedings of the Academic Track at State of the Map 2022*, Florence, Italy, 19-21 August, 2022.
[18] Nakacwa, S., McCusker, B., Sill, K., Clark, C., & Nealon, R. (2022). Combining Volunteered Geographic Information and WPdx standards to Improve Mapping of Rural Water Infrastructure in Uganda. In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.) *Proceedings of the Academic Track at State of the Map 2022*, Florence, Italy, 19-21 August, 2022.
[19] Dsoua, A., Schott, M., & Lautenbach, S. (2022). Comparative Integration Potential Analyses of OSM and Wikidata – The Case Study of Railway Stations. In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.) *Proceedings of the Academic Track at State of the Map 2022*, Florence, Italy, 19-21 August, 2022.
[20] Herfort, B., Lautenbach, S., Porto de Albuquerque, J., Anderson, J., & Zipf, A. (2022). Inequalities in the completeness of OpenStreetMap buildings in urban centers. In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.) *Proceedings of the Academic Track at State of the Map 2022*, Florence, Italy, 19-21 August, 2022.
[21] Camboim, S.P., Paiva, C., Schmidt, M., Elias, E., Bortolini, E., & Neivas, G. (2022). The cell size issue in OpenStreetMap data quality parameter analyses: an interpolation-based approach. In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.) *Proceedings of the Academic Track at State of the Map 2022*, Florence, Italy, 19-21 August, 2022.
[22] Dickinson, C., Patel, J., & Sarkar, D. (2022). Corporate editing and its impact on network navigability within OpenStreetMap. In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.) *Proceedings of the Academic Track at State of the Map 2022*, Florence, Italy, 19-21 August, 2022.
[23] Anderson, J., & Omidire, T.W. (2022). Increasing OSM Data Accessibility with the Analysis-Ready Daylight Distribution of OSM: Demonstration of Cloud-Based Assessments of Global Building Completeness. In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.) *Proceedings of the Academic Track at State of the Map 2022*, Florence, Italy, 19-21 August, 2022.
[24] Vestena, K., Camboim, S., & Santos, D. (2022). OSM Sidewalkreator - A QGIS plugin for automated sidewalk drawing for OSM. In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.) *Proceedings of the Academic Track at State of the Map 2022*, Florence, Italy, 19-21 August, 2022.
[25] Schott, M., Zell, A., Lautenbach, S., Demir, B., & Zipf, A. (2022). Returning the favor - Leveraging quality insights of OpenStreetMap-based land-use/land-cover multi-label modeling to the community. In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.) *Proceedings of the Academic Track at State of the Map 2022*, Florence, Italy, 19-21 August, 2022.
[26] Grinberger, A.Y., Schott, M., Raifer, M., & Zipf, A. (2021). An analysis of the spatial and temporal distribution of large-scale data production events in OpenStreetMap. *Transactions in GIS, 25*(2), 622–641.
[27] Lan, D., Dalyot, S. & Baram-Tsabari, A. (2022). Landmarks for accessible space – promoting geo-literacy through geospatial citizen science. In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.) *Proceedings of the Academic Track at State of the Map 2022*, Florence, Italy, 19-21 August, 2022.
[28] Cohen, A., Natapo, A., & Dalyot, S. (2022). Leveraging OpenStreetMap to investigate urban accessibility and safety of visually impaired pedestrians. In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.) *Proceedings of the Academic Track at State of the Map 2022*, Florence, Italy, 19-21 August, 2022.
[29] Chan, C.Y.-C., Weigand, M., Alnajar, E.A., & Taubenböck, H. (2022). Investigating the capability of UAV imagery for AI-assisted mapping of Refugee Camps in East Africa. In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.) *Proceedings of the Academic Track at State of the Map 2022*, Florence, Italy, 19-21 August, 2022.

[30] Solís, P. (2022). YouthMappers: A Hybrid Movement Design for the OpenStreetMap Community of Communities. In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.) *Proceedings of the Academic Track at State of the Map 2022*, Florence, Italy, 19-21 August, 2022.
[31] Steele, A.L. (2022). Mapping crises, communities and capitalism on OpenStreetMap: situating humanitarian mapping in the (open source) mapping supply chain. In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.) *Proceedings of the Academic Track at State of the Map 2022*, Florence, Italy, 19-21 August, 2022.
[32] Lin, Y.-W. (2010). A qualitative enquiry into OpenStreetMap making. *New Review of Hypermedia and Multimedia, 17*(1), 53–71.
[33] Shrestha, A., Raj Budhathoki, N., & Erbstein, N. (2022). OpenStreetMap as a tool for skill building. In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.) *Proceedings of the Academic Track at State of the Map 2022*, Florence, Italy, 19-21 August, 2022.
[34] Juhász, L., & Mooney, P. (2022). Null Island - a node of contention in OpenStreetMap. In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.) *Proceedings of the Academic Track at State of the Map 2022*, Florence, Italy, 19-21 August, 2022.
[35] Sieber, R.E., & Haklay, M. (2015). The epistemology(s) of volunteered geographic information: a critique. In: *Geo: Geography and Environment, 2*(2), 122–136.
[36] Grinberger, A.Y., Minghini, M., Juhász, L., Yeboah, G., & Mooney, P. (2022). OSM Science - the academic study of the OpenStreetMap project, data, contributions, community, and applications. In: *ISPRS International Journal of Geo-Information, 11*(4), 230.

# YouthMappers: A Hybrid Movement Design for the OpenStreetMap Community of Communities

Patricia Solís[1,]*

[1] YouthMappers, Knowledge Exchange for Resilience, School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, Arizona, United States; patricia.solis@asu.edu

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2022 Conference after peer-review.

Increasingly ubiquitous open spatial technologies offer the opportunity for new actors to participate in creating knowledge about the places where they live and work, and where they connect to others around the world. The situated nature of data production processes affects mappers and mapping, is generally understudied despite its potential to make meaningful contributions to purposeful continuously improved design of such processes for an open community like OpenStreetMap (OSM). University students are one set of actors who have grown significantly in their visibility and contributions to OSM, in part through the establishment of YouthMappers in 2015. This inclusive international network of university-based, youth-led, faculty-mentored chapters currently on more than 320 campuses in 67+ countries [1] works to mobilize and support university student mapping action that responds to humanitarian and development needs by creating and using an ecosystem of data and tools centered on OSM. The YouthMappers experience lends itself to explore interesting questions about the cultural and organizational aspects of data production and usage practices in OSM, in order to improve them. In this case, we explore these aspects as they occur within and through the academic sector, particularly through the hands and eyes of student youth. As a consortium design, this networked set of local groups works on the one hand, to create and use data on their local campuses and home communities, and on the other, to remotely contribute data on imagery-visible features in response to humanitarian, development, and knowledge needs wherever they may occur around the globe. Furthermore, they act not only within the OSM "community of communities" framework [2], but also within an existing global infrastructure of academic institutions with its own set of shared educational aims, knowledge generating practices, and cultural norms. Meanwhile, students are motivated both by learning and using new skills and workforce competencies as well as by the opportunity to participate in the world's largest volunteered geographic information project and the activities that make common good use of the data. So how do YouthMappers navigate these different aims within these different spaces of action?

To address this question, two aspects of this experience are the focus of attention in this study. First, this study aims to identify what are some of the qualitative and quantitative characteristics distinguishing the performance of YouthMappers as an academic-based

Solís, P. (2022). YouthMappers: A Hybrid Movement Design for the OpenStreetMap Community of Communities.
In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.). Proceedings of the Academic Track at State of the Map 2022, Florence, Italy, 19-21 August 2022. Available at https://zenodo.org/communities/sotm-22
DOI: 10.5281/zenodo.7004448

 

community within OSM. Second, this study aims to better understand how the design approach taken by and on behalf of YouthMappers reinforces an identity as unique contributors. The design of the YouthMappers consortium within the context of OSM is both purposeful and justified by the literature [3,4]. Our approach builds upon our typology of understanding the internal OSM community dynamic as a multi-scaled, social ecosystem, which can be categorized across non-mutually exclusive groups using typical descriptors such as sector-based characteristics; modality of engagement with OSM (i.e. editors; map-based services providers; data consumers); and around social bonds formed for purposes (e.g. HOT) or identity (e.g. GeoChicas), or place (e.g. local OSM communities), as well as other dimensions of connectedness. The case of YouthMappers demonstrates a multi-scaled social ecosystem that could be tagged as public/non-profit, academic, editor-consumer, purpose-driven, identity and place-based [4].

Study results should be contextualized with an understanding of the current state of higher education, particularly the present tension around higher ed institutions' purpose as sites for both workforce preparation and global citizenship [5]. YouthMappers members sector, role, purpose, and identity are integral to the design of the rewards and motivations for participation. Trends in higher education, especially this tension, have been integral to the design and engagement characteristics of YouthMappers. For example, both formal credit-bearing or credential-enhancing experiences and resources are combined with the fieldwork and social group leadership model of youth-led chapters. The latter point on youth leadership is situated with reference to scholarship on the global targets of the Sustainable Development Goals (SDGs), as the currently predominant discourse for international action across humanitarian domains and within the academic sector alike. This review sets up three interlocking hypotheses we anticipate can be rejected: H1: (Action-of-Performance) Participating youth either map only locally or remotely, but not both; H2: (Hybrid-Roles) Participating youth cannot simultaneously pursue personal aims to prepare themselves for the workforce and to express their identities as global citizens; and H3: (Movement-Minded) Participating youth cannot articulate the impacts/benefits of actions undertaken for broader communities or society through their work with OSM, nor identify the roles/contributions of youth action in this work for the common good. Data to test the first hypothesis relate to performance and include a range of metrics of participation [6–8], statistics of known users [9], and a review of data from other studies of YouthMappers' editing contributions [10]. To track and monitor the mapping activity and growth of YouthMappers, we analyzed the OSM database to identify contributions from usernames provided to the network as being associated with a particular YouthMappers chapter in annual reports and queries [11]. This contributor-centric approach is not comprehensive, but considered to more closely represent the active community than a hashtag-based approach, as the database contains a sufficiently large estimated sample of the YouthMappers population to provide pattern analysis on the quantity and types of contributions they make to OSM, where response rates are estimated at 78% [11]. Local mapping is defined as any edit made within the same country as the location of the chapter for which each username is identified; remote mapping is defined as edits in any other country than where their home chapter is based. Edits were tabulated by building, road, amenity, and all features. Gender is known for approximately 41% of usernames and a descriptive analysis of performance shows different patterns of editing. YouthMappers whose edits are exclusively local represent 29% of the community; whose edits are exclusively remote only 23% of the community; and nearly half

(48%) map both local and remote. Output by tendency to edit however is greatly skewed towards the half who map both locally and remotely, a subgroup which accounts for fully 96% of the edits, compared to only 1% and 3% respectively for exclusively local and exclusively remote mappers. When removing outliers (defined as the 8 super mappers who produce more than three times the standard deviation from the median), this pattern persists, whereby YouthMappers who map both locally and remotely produce 91% of edits, compared to only 3% and 6% of exclusively local and exclusively remote mappers. By edit type, 60.4% of buildings are mapped locally, 56.6% of highways and 94.9% of amenities are mapped by YouthMappers from the same country as where the feature appears. Gender-based comparisons reveal that women mappers tend to be slightly more likely to map remotely, least likely to map highways locally and map amenities at the same rates.

Data to test the second hypotheses relate to identity and come from the long-running student-authored blogs [12], as well as a global survey of YouthMappers collected in 2019 accompanied by a qualitative set of member queries to iterate interpretation of the survey results [11,13]. YouthMappers blogs are student-authored and available for all members as an opportunity to vocalize insights about mapping experiences and perspectives. Two years of blog narratives were reviewed [12] and used as reflective or open-ended writing that education literature validates for textual analysis to explore both authentic task and qualitative assessment via informal writing that is coded by independent multiple readers who are neither authors nor program managers using both a deductive and an inductive (emergent) content analysis framework [12]. These blogs were also categorized by the region of the author and by gender and included only blogs from students who spoke to experiences, excluding announcements or events. All 302 utterances were coded and analyzed based upon the locus of affect, that is being from an individual or a group or collective perspective. Results demonstrated that regardless of region, students tended to discuss positive aspects of learning as members of a group as well as learning as individuals, where positive affect occurred one-third to one-fourth more frequently in the former category. These utterances point to ease in navigating the tension of individual benefits of participation (e.g. preparing oneself for work) while at the same time enjoying group benefits (e.g. being part of a collective of citizen action). Between 20 to 30% of respondents reported receiving a paid internship as a direct result of YouthMappers participation, and for more than 10% of them, doing so led to a job offer [11].

Data to test the third hypothesis come from the same sources and a set of case studies on the SDGs [13]. In particular, our surveys reveal across all regions, across gender, across various lengths of participation time, that students reflect above 90 % that they value YouthMappers participation because of motivation for being a good citizen, social responsibility, living a well-rounded life, as well as for finding a well-paying job and finding a rewarding career [11]. Some differences are perceived by region of origin. YouthMappers from the Global North tended to agree more strongly with knowing how to explain benefits of science to society, while YouthMappers from the Global South were statistically significantly more likely to know how mapping could impact their local community [11].

These results reveal a spectrum of interests balancing local and global mapping across the consortium, and across regions, and other axes of participation. They also indicate the extent to which youth reflect on local benefits, including personal skill development, versus global citizenship, including how they understand the meanings of their actions for SDGs, locally and globally. Detected differences by gender, world region, and duration of participation are interpreted and validated with additional qualitative data. The results confirm that we may reject all three hypotheses, validating the model. Findings build a case for understanding the potential of community design to advance the goals of OSM broadly. We also point to the role of university systems as third space sites for enabling performance and identities for youth action [5,14]. This means universities can serve as hybrid organizations that, when linked to global discourses on issues like sustainability (SDGs) and open data (OSM), can mobilize youth to create and participate in "hybrid movements" [5] that are both frameworks of performance and spaces of identity to advance OSM communities. Our findings offer insights for how other types of communities could leverage their existing milieu in ways that strengthens OSM broadly. The details and configuration need to be designed according to the framework of what type of community and what context and institutional milieu they are situated within [2,4]. Ultimately, the idea of hybrid movements encourages OSM to embrace a pluralistic, inclusive and diverse set of communities that not only bring individual contributions but leverages other systems like the landscape of academia was systematically enlisted via the YouthMappers design.

**References**

[1] YouthMappers (2022). Chapter Map. Retrieved from https://www.youthmappers.org/chapters
[2] Solís, P. (2016). YouthMappers: Open Mapping for Resilience. Presented at HOT Summit, Brussels, Belgium, September 22, 2016.
[3] Solís, P., McCusker, B., Menkiti, N., Cowan, N., & Blevins, C. (2018). Engaging Global Youth in Participatory Spatial Data Creation for the UN SDGs. *Applied Geography, 98*, 143–155.
[4] Brovelli, M.A., Ponte, M., Schade, S., & Solís, P. (2019). Citizen Science in Support of Digital Earth. In: Guo, H., Goodchild, M.F., & Annoni, A. (Eds.) *Manual of Digital Earth*, Springer, Singapore, 593–622.
[5] Solís, P., Rajagopalan, S., Villa, L., Mohiuddin, M.B., Boateng, B., Wavamunno Nakacwa, S., & Peña Valencia, M.F. (2022). Digital Humanitarians for the Sustainable Development Goals: YouthMappers as a Hybrid Movement. *Journal of Geography in Higher Education, 46*(1), 80–100.
[6] Andal, F., Islam, M., Shaheen, A.R., & Anderson, J. (in press). Understanding YouthMappers Contributions to Building Resilient Communities in Asia. In: Solís, P., & Zeballos, M. (Eds.) *Open Mapping towards Sustainable Development Goals*, Springer, Cham.
[7] Boateng, E.N.K., Manyungwa, Z., & Anderson, J. (in press). Assessing YouthMappers Contributions to the Generation of Open Geospatial Data in Africa. In: Solís, P., & Zeballos, M. (Eds.) *Open Mapping towards Sustainable Development Goals*, Springer, Cham.
[8] Walachosky, N., Gómez, C., Martínez, K., Amaya, M., Rodríguez, M., Centeno, M., & Anderson, J. (in press). Visualizing YouthMappers Contributions in Latin America. In: Solís, P., & Zeballos, M. (Eds.) *Open Mapping towards Sustainable Development Goals*, Springer, Cham.
[9] Anderson, J. (2022). YouthMappers Activity. Retrieved from https://activity.youthmappers.org
[10] Mahmud, Z., Chauhan, A., Sarkar, D., & Soden, R. (2022). Revisiting Engagement in Humanitarian Mapping. In: *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*, New Orleans, LA, USA, 29 April-5 May 2022.
[11] Solís, P., Anderson, J., & Rajagopalan, S. (2020). Open Geospatial Tools for Humanitarian Data Creation, Analysis, and Learning through the Global Lens of YouthMappers. *Journal of Geographical Systems*, *23*(4), 599–625.
[12] Hite, R., Solís, P., Wargo, L., & Larsen, T.B. (2018). Exploring Affective Dimensions of Authentic Geographic Education using YouthMappers Blogs. *Education Sciences, 8*(4), 173.
[13] Solís, P., & Zeballos, M. (Eds.) (in press). *Open Mapping towards Sustainable Development Goals: Voices of YouthMappers on Community Engaged Scholarship*. Springer, Cham.
[14] Soja, E. (1996). *Thirdspace: Journeys to Los Angeles and Other Real-and-Imagined Places*. Blackwell, Oxford.

# Mapping crises, communities and capitalism on OpenStreetMap: situating humanitarian mapping in the (open source) mapping supply chain

Anne Lee Steele[1,*]

[1] Anthropology and Sociology, The Graduate Institute of International and Development Studies, Geneva, Switzerland; anne.steele@graduateinstitute.ch

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2022 Conference after peer-review.

Founded in the aftermath of the 2010 earthquake in Haiti, the Humanitarian OpenStreetMap Team (HOT) helps both globally remote and local in-person volunteers to identify roads, buildings, and other features on the OpenStreetMap (OSM) platform. Created as a "free, editable map of the world," OSM has enabled the mass-creation of volunteered geographical information (VGI) on a scale that is now more accurate than proprietary maps in many places, particularly as "crisis-mapping" has emerged as a means to gather real-time data on areas that have been affected by natural disasters or socio-political conflicts. OSM has also become a site of resistance, where local and indigenous communities have engaged in mapping projects to reclaim autonomy, agency, and space through the historically contested practice of (digital) mapping. For these reasons, such crowdsourced maps have increasingly been used by humanitarian organisations to facilitate aid and disaster relief, and as open training data for algorithms learning how to automatically detect features through Artificial Intelligence (AI). As a key partner of humanitarian, corporate, and local actors, and having mobilised over 200,000 volunteers since 2010, HOT lies at the crux of these ongoing entanglements and contestations, both within and around the field of OSM.

Previous studies of crowdsourced geographical information and crisis-mapping have generally revolved around quantitative analyses of OSM's data, focusing on the credibility of the data itself, the makeup of the communities that contribute to it, the effects of "event-centric" crowdsourcing, or "newcomer retention" in humanitarian mapping [1]. Alternatively, they have also focused on the "spatial knowledge", "hacker political imaginary", and gender composition of mappers themselves [2].

Parallel studies of other volunteer-driven communities like "Wikipedians" have taken similar approaches, analysing "user-generated content" and the motivations behind them [3]. Both hacking [4] and free and open source software (F/OSS) [5] have also been explored ethnographically. While automated detection of features on OSM has only recently become

Steele, A.L. (2022). Mapping crises, communities and capitalism on OpenStreetMap: situating humanitarian mapping in the (open source) mapping supply chain.
In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.). Proceedings of the Academic Track at State of the Map 2022, Florence, Italy, 19-21 August 2022. Available at https://zenodo.org/communities/sotm-22
DOI: 10.5281/zenodo.7004452

 

an important topic of research, ongoing studies have primarily focused on the accuracy or credibility of this endeavour [6].

While existing studies of digital communities have focused on the socialities they engender or labor they require, they tend to forget the bureaucratic apparatuses that have emerged to govern them, both implicitly and explicitly. Similarly, studies of humanitarianism have focused on the ethics they operationalize, or the technologies that are mobilized in turn, but often at the expense of engaging in the wider spectrum of social and economic life that they enable. [7–10] While this project draws upon these overlapping strains of research, it seeks to push the debate in an ethnographic direction, scaffolded by theories of bureaucratic technology, political economy, and humanitarianism.

This research draws from participation in over 40 online events over 1.5 years, including mapathons, conferences and online lectures with OSM mappers, as well as semi-structured interviews conducted with 27 key-informants, alongside watching more conference videos, and reading blogs, mailing list emails, Twitter exchanges, and other internet archives. While empirically influenced by studies of hacking and open source software, this work ultimately focuses on the mechanisms and means through which this "free and open map" is created, and ultimately the ways of seeing and doing that it enables.

The "supply chains" heuristic emerged as a means to understand and illustrate this process. Similar to how supply chains "link ostensibly independent entrepreneurs, making it possible for commodity processes to span the globe", the OSM project relies upon a series of interconnected processes that enable the creation of the world's crowdsourced map [11]. Similar to how the satellite, computer, and software industries converged to create the conditions that allowed for OSM's creation, so do people – and their associated institutions create map data through an almost miraculous collision of circumstances, assured precedent, and training. HOT, which is the initial entry point into open source mapping, made its name by optimizing the mapping value chain: that is, by making it easier to contribute to OSM. But it also extended outwards: contributing to OSM was enabled not only by the wider socio-economic forces that coalesced to produce the project in the first place, but also by a series of digital value chains – both past and present.

By delineating this supply-chains approach, this study hopes to scaffold a mental model of humanitarian mapping and the OSM more broadly, to be employed in future studies – both quantitative and qualitative. Practically, it hopes to provide a heuristic and application of ethnographic tools, and present questions and queries directly to the community more broadly.

**References**

[1] Dittus, M., Quattrone, G., & Capra, L. (2016). Analysing volunteer engagement in humanitarian mapping: Building contributor communities at large scale. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 27 February 2016, 108–118.

[2] Brandusescu, A., & Sieber, R.E. (2018). The spatial knowledge politics of crisis mapping for community development. *GeoJournal, 83*(3), 509–524.

[3] Graham, M., Straumann, R.K., & Hogan, B. (2015). Digital Divisions of Labor and Informational Magnetism: Mapping Participation in Wikipedia. *Annals of the Association of American Geographers, 105*(6), 1158–1178.

[4] Kelty, C. (2008). *Two Bits: The Cultural Significance of Free Software*. Duke University Press, Durham, NC.

[5] Coleman, G. (2012). *Coding Freedom: The Ethics and Aesthetics of Hacking*. Princeton University Press, Princeton, NJ.
[6] Brovelli, M.A., Minghini, M., Molinari, M., & Mooney, P. (2017). Towards an Automated Comparison of OpenStreetMap with Authoritative Road Datasets. *Transactions in GIS, 21*(2), 191–206.
[7] Cross, J. (2020). Capturing Crisis. *The Cambridge Journal of Anthropology, 38*(2), 105–124.
[8] Redfield, P. (2016). Fluid technologies: The Bush Pump, the LifeStraw® and microworlds of humanitarian design. *Social Studies of Science, 46*(2), 159–183.
[9] Scott-Smith, T. (2016). Humanitarian neophilia: the 'innovation turn' and its implications. *Third World Quarterly, 37*(12), 2229–2251.
[10] Ticktin, M. (2014). Transnational humanitarianism. *Annual Review of Anthropology, 43*, 273–289.
[11] Tsing, A. (2009). Supply chains and the human condition. *Rethinking Marxism*, *21*(2), 148–176.

# Automated derivation of public urban green spaces via activity-related barriers using OpenStreetMap

Theodor Rieche[1, 2,*] and Robert Hecht[1]

[1] Leibniz Institute of Ecological Urban and Regional Development, Dresden, Germany; t.rieche@ioer.de, r.hecht@ioer.de
[2] University of Applied Sciences (HTW), Dresden, Germany

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2022 Conference after peer-review.

In addition to important ecosystem services such as clean air or local climate regulation, green spaces also provide tranquility and recreational services, contributing to a good quality of life for the population. In high-density urban areas, publicly accessible green spaces are used for a variety of recreational activities, which has become even more important, not least because of the COVID-19 pandemic [1–4]. In this context, the research project "Information and Navigation on Urban Green Spaces in Cities - meinGruen" examined publicly accessible green spaces with regard to a variety of criteria in order to assess their suitability for the pursuit of leisure activities, such as going for a walk or playing soccer [3,5]. The aim of this study is to derive a suitable polygon dataset to describe the spatial distribution of publicly accessible urban green spaces. The presented approach favors the use of OpenStreetMap (OSM) data and intrinsic knowledge. Advantages of the use of OSM data are the global availability, the often high completeness in urban areas as well as the unified open data license ODbL 1.0. In this way, problems with data availability and heterogeneity due to different responsible authorities can be avoided. Ludwig et al. [6] describe an approach to mapping public green spaces based on OSM and Sentinel-2 satellite imagery in which barriers and land use changes are considered based on a priori (expert knowledge) assumptions for polygon generation. In the approach presented here, spatial delimitation is to be refined by describing barriers by probability values. The term "barrier" is first analyzed in an interdisciplinary way in order to then work out its meaning for the spatial delimitation of a green space. Here, barriers delimit the action space of a recreational activity. They represent a spatial boundary in the real world over which the respective activity cannot be performed. While for a number of linear object types (such as walls, fences, rivers, roads or railroad lines) it can be assumed for certain that they are barriers, there are others (such as paths or the change of land use) for which knowledge is still lacking. The approach goes beyond the barriers modeled in OSM (barrier=*) by including trails and land use changes, among others. The generated spatial delineation of green spaces should thus be

Rieche, T., & Hecht, R. (2022). Automated derivation of public urban green spaces via activity-related barriers using OpenStreetMap.
In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.). Proceedings of the Academic Track at State of the Map 2022, Florence, Italy, 19-21 August 2022. Available at https://zenodo.org/communities/sotm-22
DOI: 10.5281/zenodo.7004458

 

more realistically adaptable to the needs of users. The study area includes the city of Dresden in Germany, plus a buffer of five kilometers. OSM represents the main data source. For training and validation, official cadastral data (ALKIS) as well as a dataset on cadastral parcels owned by the city of Dresden were used.

The methodology consists of six steps: First, according to defined rules, types of barriers were extracted from OSM data. Second, we derived a land use layer without overlaps and holes from OSM. Here, two options were compared regarding different target schemas for land use classification. Third, a mapping in terms of a “ground-truth“ in selected areas in Dresden followed in order to be able to evaluate the existence of a barrier on site for the extracted paths and changes of land use. Fourth, generic probabilities for the existence of a barrier were determined based on path type or land use change type. Fifth, a polygon mesh was created by applying thresholds to the determined barrier probabilities. Sixth, the generated polygons were enriched with attributes on the number of green space-related POI, such as benches, trash cans, or trees. Models for "greenness" and "accessibility" are thereby trained. For the technical implementation mainly Docker, PostgreSQL/PostGIS, Python (Geopandas, Scikit-Learn) and Jupyter Notebook were used. Data import was performed by osm2pgsql and ogr2ogr. For the on-site mapping, we used the app QField.

Land use layers were successfully generated for the study area using a residual class. The results indicated that the land use classification according to the area schema of the IOER-Monitor (option B) has a higher thematic accuracy with a maximum of 33 classes (433 original OSM tags were assigned) than the option A based on a classification according to osmlanduse.org, see Schultz et al. (up to 13 classes, based on 61 OSM tags) [7,8]. The classes of arable land (A: 28.40% / B: 28.06% share of area) as well as forest (A: 21.81% / B: 23.33%) are dominant in both variants. While the residual class takes up 6.29% of the area in option A, it is only 4.88% in option B. For the “ground-truth”, a total of approximately 82.3 km of paths (with 408 line objects) and approximately 64.2 km of land use changes (1720 line objects) were evaluated for the presence of a barrier in two selected areas in Dresden. The land use changes are based on variant B. Data were collected on 61 different land use transitions and four different trail types. While bike lanes can be safely assumed to be barriers (100.0%), the "track" (96.8%), "footway" (92.7%), and "path" (86.0%) trail types have slightly lower barrier probabilities. Among land use transitions, the forest-meadow (12.6%), meadow-sports facility (22.8%), meadow-park (24.6%), and forest-grassland (26.7%) transitions have the lowest barrier probabilities. Together with the barriers assumed to be safe at the beginning, a line pool is formed, from which different polygon meshes are generated based on different intervals for the barrier probability ($p \geq 0\%$; $p \geq 20\%$; $p \geq 40\%$; $p \geq 60\%$; $p \geq 80\%$; $p = 100\%$). The lower the probability threshold, the higher the number of polygons created and the smaller their area becomes. For the "accessibility" model, the number of benches, trash cans, public toilets and public internet were considered per polygon. The application of a logistic regression model achieved 76.7% accuracy, with similar values using a support vector classifier (76.8%). The "greenness" model is based on the number of benches, picnic tables, trees, and trash cans per polygon. The accuracy is about 92.2% using a logistic regression model and 92.3% using a support vector classifier.

This work successfully demonstrates a new approach to derive publicly accessible urban green spaces based on OSM data considering different qualities of barriers. Based on the examined barrier probability of path types and land use transitions, more realistic spatial delineations of green spaces were made possible. The chosen approach is globally applicable due to the use of OSM. In each case, locally prevailing climatic and cultural influences must be taken into account. The knowledge collected here can be applied in the Central European region. For other areas, a renewed “ground-truth” may have to be carried out on site. The schematic transformation of the land use into the area schema of the IOER-Monitor leads to a reduction of classes compared to the original data. In addition to benefits in capturing barrier probability, it also simplifies comparability and transferability. Thus, other data could also be migrated into this schema. The polygon generation based on different barrier probabilities allows here a differentiated setting of the desired action space for the relevant leisure activities. The quality of the trained models is good, but can be improved. In order to improve the accuracy a variety of additional features could be calculated for each potential green space polygon, such as path network density or density of path network intersections (see also Ludwig et al. [6]) and considered in the model. Questions about the perception and use of green spaces can also be part of future interdisciplinary research. The tested approach could be adapted to validate the completeness of mapped fences, walls or hedges in OSM. An automatically processed layer could also be a motivation for the OSM community to map gaps in land use. Furthermore, the knowledge of barrier probabilities and accessibility could support algorithms in the field of open space routing to determine a more realistic shortest path through publicly accessible spaces [9]. Thus, more accurate analysis for green space supply and green space accessibility for the population become possible.

## References

[1] Bolund, P., & Hunhammar, S. (1999). Ecosystem services in urban areas. *Ecological Economics, 29*(2), 293–301.

[2] Perrings, C., Naeem, S., Ahrestani, F., Bunker, D.E., Burkill, P., Canziani, G., Elmqvist, T., Ferrati, R., Fuhrman, J., Jaksic, F., Kawabata, Z., Kinzig, A., Mace, G.M., Milano, F., Mooney, H., Prieur-Richard, A.-H., Tschirhart, J., & Weisser, W. (2010). Ecosystem Services for 2020. *Science, 330*(6002), 323–324.

[3] Krellenberg, K., Artmann, M., Stanley, C., & Hecht, R. (2021). What to do in, and what to expect from, urban green spaces – Indicator-based approach to assess cultural ecosystem services. *Urban Forestry & Urban Greening, 59*, 126986.

[4] Ugolini, F., Massetti, L., Calaza-Martínez, P., Cariñanos, P., Dobbs, C., Ostoić, S.K., Marin, A.M., Pearlmutter, D., Saaroni, H., Šaulienė, I., Simoneti, M., Verlič, A., Vuletić, D., & Sanesi, G. (2020). Effects of the COVID-19 pandemic on the use and perceptions of urban green space: An international exploratory study. *Urban Forestry & Urban Greening, 56*, 126888.

[5] Hecht, R., Artmann, M., Brzoska, P., Burghardt, D., Cakir, S., Dunkel, A., Gröbe, M., Gugulica, M., Krellenberg, K., Kreutzarek, N., Lautenbach, S., Ludwig, C., Lümkemann, D., Meinel, G., Schorcht, M., Sonnenbichler, A., Stanley, C., Tenikl, J., Wurm, M., & Zipf, A. (2021). A web app to generate and disseminate new knowledge on urban green space qualities and their accessibility. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, VIII-4/W1-2021*, 65–72.

[6] Ludwig, C., Hecht, R., Lautenbach, S., Schorcht, M., & Zipf, A. (2021). Mapping Public Urban Green Spaces Based on OpenStreetMap and Sentinel-2 Imagery Using Belief Functions. *ISPRS International Journal of Geo-Information, 10*(4), 251.

[7] IÖR Monitor (2022). Methodology—Monitor of Settlement and Open Space Development (IOER Monitor), Land Use Typology. Retrieved from https://www.ioer-monitor.de/en/methodology

[8] Schultz, M., Voss, J., Auer, M., Carter, S., & Zipf, A. (2017). Open land cover from OpenStreetMap and remote sensing. *International Journal of Applied Earth Observation and Geoinformation, 63*, 206–213.

[9] Hahmann, S., Miksch, J., Resch, B., Lauer, J., & Zipf, A. (2018). Routing through open spaces – A performance comparison of algorithms. *Geo-Spatial Information Science, 21*(3), 247–256.

# Landmarks for accessible space – promoting geo-literacy through geospatial citizen science

Dalit Lan[1], Ayelet Baram-Tsabari[1] and Sagi Dalyot[2,*]

[1] Education in Science and Technology, The Technion, Haifa, Israel; lan.dalit@campus.technion.ac.il, ayelet@technion.ac.il
[2] Mapping and Geoinformation Engineering, The Technion, Haifa, Israel; dalyot@technion.ac.il

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2022 Conference after peer-review.

The 21$^{st}$ Century dictates that people have a good spatial and geographic understanding and knowledge. Geographic literacy, or geo-literacy, is aimed to provide skills to read, interpret and use geospatial information [1]. This is achieved by acquiring critical spatial thinking, reasoning, and analysis, and presenting understanding of the world using geographical terms and spatial language [2]. Recent years have led to the development of new geo-literacy education programs specifically designed to nurture and promote these skills [3]. These education programs build on geographic education that promotes spatial thinking and active citizenship [4]. Still, little evidence exists regarding the potential and capacity of these programs in advancing civic and geographic skills and knowledge in the 21$^{st}$ Century, and on its contribution to – and advancing of – the individual and the society.

The aim of this research is to gain a better understanding of the development of geo-literacy in the framework of a citizen science project in high schools. The citizen science project implemented in several high schools in Israel – landmarks for accessible space, advances scientific research that aims to make the urban environment more accessible for visually impaired pedestrians. The participating high school students practice participatory mapping with OpenStreetMap (OSM) to map features relevant to the navigation of visually impaired pedestrians. These map features are used for the automatic calculation of optimal walking routes [5]. The project combines social involvement, learning through geographic information systems, and familiarity with the field of urban accessibility for visually impaired people.

The project includes the following stages:

1. Pre-stage that includes a) the design of the modular learning environment, b) the organizational and pedagogical preparation of the project integration in schools, and c) questionnaires examining the current level of geographic literacy of the participating students and their perspectives regarding the integration of citizen science in schools.

Lan, D., Baram-Tsabari, A., & Dalyot, S. (2022). Landmarks for accessible space – promoting geo-literacy through geospatial citizen science.
In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.). Proceedings of the Academic Track at State of the Map 2022, Florence, Italy, 19-21 August 2022. Available at https://zenodo.org/communities/sotm-22
DOI: 10.5281/zenodo.7004476

 

2. Intervention program that includes guest seminars (including YouTube videos), lectures and learning activities, exposure to the world of visually impaired people, and the need for accessible environments and learning activities in the field of geoinformation with emphasis on OSM, crowdsourcing, and participatory mapping.
3. Mapping missing data into OSM. This stage is carried out in the field with a designated app developed for this project. The app – "Mundi", depicted in Figure 1 – allows the mapping of specific geographic features (mapping elements) used for the calculation of accessible routes designed specifically for visually impaired pedestrians. The features include, among others, sidewalks, crossings, accessibility aids, and handrails. The app includes gamification and tasks to encourage the students to map the missing features in their area of residence.
4. Post-stage questionnaires, aimed to investigate and analyze the development of spatial skills in the context of participation in this program, examining whether students' level of geographic literacy improved and whether they gained new knowledge on urban accessibility and the navigation proficiencies of visually impaired pedestrians. This stage also included a quantitative analysis of the students' contributions in terms of OSM mapping, among others, the number of map edits, type of mapped features, the spatial coverage and temporal extent of their mapping activity.

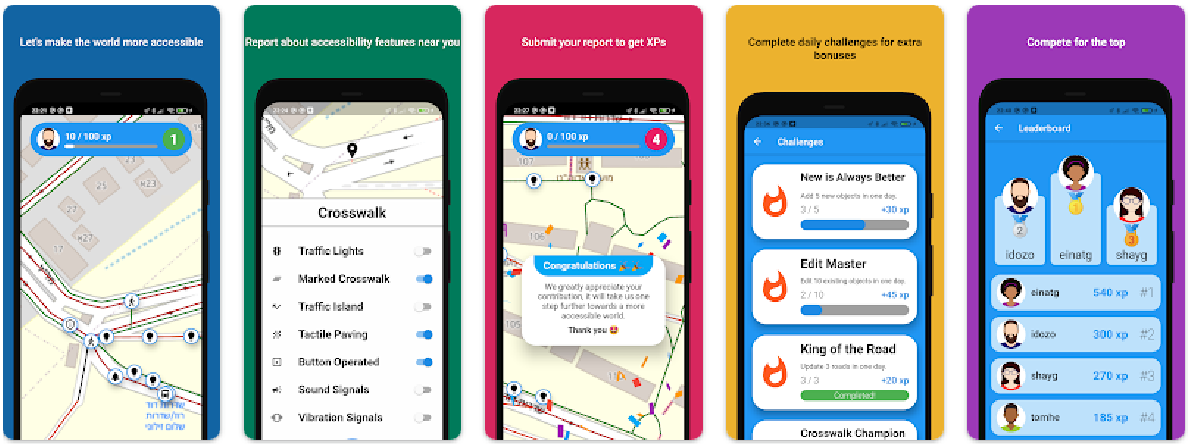


*Figure 1*. Mundi app screens (from left to right): OSM map with existing map features, updating attributes of a crosswalk feature, submission of a new feature, daily challenges, and leaderboard (source: https://play.google.com/store/apps/details?id=com.technion.access_map).

The study was conducted in the last two years in 13 high schools, including 25 classes and 460 students. The intervention model was implemented for three months in each class. The participating students implemented this project within their Cyber Geography studies, enabling them to learn through various geographic information system projects. Based on the analysis of 5 high schools who mapped with the Mundi app, close to 5,000 OSM edits were made by the students, which included more than 600 new crossings that include more than 1,000 attributed tags, more than 100 new sidewalks (as way attribute) with attributes (e.g., shared lane with bicycles), and more than 3,000 new street objects and obstacles (e.g., bus stations, light poles, trees, gates, bicycle parking).

Preliminary analysis showed that participation in the citizen science project increased the students' geospatial thinking and reasoning. For example, according to the questionnaire variables, on a score scale of 0-100, the geospatial thinking score has increased from 31 to 56, while the spatial awareness score has increased from 34 to 73 (paired samples t-test, $p < .001$). The geographic skills knowledge has increased from 3 to 3.9 (scale of 1-5). Moreover, the students' self-esteem with respect to their knowledge and use of geographic skills has improved considerably - from 16 to 56. In addition, results show that the broad and in-depth intervention model increased the students' appreciation of the scientists' contribution to the project, the contribution of the program in general, and the satisfaction with their participation in the project.

Current research recognizes the potential of implementing geospatial citizen science projects that introduce participatory mapping in schools [6]. The contribution of GIS in general - and OSM in particular - to citizen science projects promotes many challenges, as well as contributions. For example, examining the motives of geographic information volunteers, strengthening the motivation to volunteer over time, developing spatial literacy skills, and improving the quality and accuracy of the data [7]. This is the first project to introduce the use of OSM-based learning to the study of geography in Israel. Based on its outcome and analysis, we believe that this research contributes to various pedagogic and education levels, in terms of theoretical knowledge about the integration of innovative geo-literacy programs. These promote the drawing of operational and applicable actions regarding the planning of future projects and serving stakeholders in academia and the education system in terms of integrating scientific projects that increase students' involvement in science and society and promote geo-literacy.

**References**

[1] Turner, S., & Leydon, J. (2012). Improving geographic literacy among first-year undergraduate students: Testing the effectiveness of online quizzes. *Journal of Geography, 111*(2), 54–66.
[2] Edelson, D.C. (2011). Geo-Literacy: Preparation for 21st Century Decision-Making. *National Geographic*. Retrieved from https://media.nationalgeographic.org/assets/file/Geo-Literacy_Backgrounder.pdf
[3] Bednarz, S., Heffron, S., & Huynh, N. (Eds.) (2013). *Road map for 21st century geography education*. Association of American Geographers, Washington, DC.
[4] Bednarz, S.W., & Bednarz, R.S. (2015). Brave new world: Citizenship in geospatially enriched environments. *Journal for Geographic Information Science, 1-2015,* 230–240.
[5] Cohen, A., & Dalyot, S. (2021). Route planning for blind pedestrians using OpenStreetMap. *Environment and Planning B: Urban Analytics and City Science, 48*(6), 1511–1526.
[6] Hennig, S. (2018). Citizen science and OpenStreetMap: Potential and challenges. *Austrian Citizen Science Conference 2018*, 51–54.
[7] Skarlatidou, A., & Haklay, M. (Eds.) (2020). Geographic Citizen Science Design: No One Left Behind. UCL Press, London.

# Comparative Integration Potential Analyses of OSM and Wikidata – The Case Study of Railway Stations

Alishiba Dsoua[1,*], Moritz Schott[2] and Sven Lautenbach[2,3]

[1] Data Science & Intelligent Systems, University of Bonn, Bonn, Germany; dsouza@cs.uni-bonn.de
[2] Institute of Geography/GIScience, Heidelberg University, Heidelberg, Germany; moritz.schott@uni-heidelberg.de, sven.lautenbach@uni-heidelberg.de
[3] HeiGIT, Heidelberg University, Heidelberg, Germany;

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2022 Conference after peer-review.

OpenStreetMap (OSM) is one of the richest and most diverse sources of geographic information. However, it lacks a fundamental property vital for spatio-semantic analyses: hierarchical structure and semantic linkage. OSM provides links to existing knowledge graphs (structured data that conforms to a specific ontology), e.g., via the *wikidata* key. The usage of these link-tags is currently limited to a small percentage of both OSM and Wikidata objects. Efforts were undertaken to enhance geographic linking (i.e., linking nearby objects of the same type) and semantic linking [1–3]. The WorldKG knowledge graph [4] provides a semantic mapping of a large subset of OSM. While the free and open OSM tagging scheme is an integral part of the OSM project that enabled its success, WorldKG overcomes the inherent lack of structure the tagging scheme represents, paving the way for a knowledge-graph integration of the OSM dataset. Still, open knowledge graphs and OSM are not fully integrated on schema and entity level.

The following analyses provide a series of comparative data insights that help to better understand the potential and implications of integration between knowledge graphs and OSM. This work compares OSM to Wikidata, one of the largest open knowledge graph projects from the Wikimedia Foundation that provides structured storage to other Wikimedia projects such as Wikipedia. Wikidata can, in many aspects, be compared to OSM in terms of its community structure, free and open nature, and simple contribution framework. In this work, the two datasets are first compared in size, data structure, and distribution. Later, we extend our analyses with a community comparison. The analyses also examine how two separate online communities with similar interests have evolved over time.

Grasping the size of the two projects is a straightforward task and visible on their websites: OSM features around 1 billion elements [5], while Wikidata is much smaller with around 97 million objects, of which approximately 9 million have geographic coordinates. Yet, the aforementioned schematic misalignment makes a comparison on dataset level

Dsouza, A., Schott, M., & Lautenbach, S. (2022). Comparative Integration Potential Analyses of OSM and Wikidata – the Case Study of Railway Stations.
In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.). Proceedings of the Academic Track at State of the Map 2022, Florence, Italy, 19-21 August 2022. Available at https://zenodo.org/communities/sotm-22
DOI: 10.5281/zenodo.7004483

 

impractical. OSM tags allow no predefined distinction between the classes and attributes of an object. While lists exist that facilitate the distinction (e.g., Map Features https://wiki.openstreetmap.org/wiki/Map_features), it is not clear on a structural level, e.g., for an object with *highway=elevator* as a tag, the user must know the tag's meaning to extract the correct class. Wikidata classes are defined using an attribute called "instance of" (abbreviation P31). For example, a mosque in Wikidata will have the attribute "instance of religious building". In addition to the incompatible schemata, uneven distribution of classes, community interests, and priorities are some of the biggest challenges for data integration. Wikidata is not limited to geographic entities, wherefore humans (people of interest) represent the largest data type. Therefore, while the analyses at the dataset level as a whole seem possible, they can not be applied to all classes due to varied definitions, incomplete or missing representations, or lack of other comparison parameters such as geographic coordinates.

The topic of railway stations was chosen in the presented study because these objects have a comparable definition and are well represented in both datasets with ca. 130k and 100k elements in OSM and Wikidata, respectively, indicating integration potential. In OSM, railway stations are mapped by the tags *railway=station* or *railway=halt*. In Wikidata, the 'instance of Q55488 (railway station)' value represents Railway Stations.

The presented work (available under the GNU Affero General Public License v3 at https://gitlab.gistools.geog.uni-heidelberg.de/giscience/ideal-vgi/osm-wikidata-comparison) provides a set of generalizable indicators for VGI project description, comparison, and monitoring. Similar approaches have been established for OSM contributors [6], single OSM elements [7], and small geographic regions [8]. For data collection in Wikidata, Wikidata API (https://www.wikidata.org/w/api.php) and Wikidata SPARQL endpoint were used. For Wikidata objects mapped with 'Railway Station', their revision history containing user information, timestamps, and the number of properties was collected. Overall contributions were collected from all users who have contributed to at least one object typed 'Railway Station'. OSM data collection was done using the ohsome framework (https://ohsome.org) to extract all railway stations mapped in OSM, including their history and all edits made by the users who edited these railway stations. In addition to a general comparison between the datasets, we defined five subsets for a more detailed comparison: OSM railway stations with links to Wikidata (59,441 elements), OSM stations without links to Wikidata (74,659), Wikidata stations that have links from OSM (45,050), Wikidata stations without links to OSM but with geocoordinates (54,594) and Wikidata stations without links to OSM and without geocoordinates (6,714).

Our first analysis regarding the growth rate of the two sources showed that OSM is reaching a saturated state regarding the number of railway stations (see Figure 1), where relatively few stations have been added since 2016. Wikidata, on the other hand, still experiences a stable number of new stations that are added to the project. The two datasets depict no clear temporal correlation hinting towards two independent communities, meaning that additions in OSM are not followed by additions in Wikidata and vice versa. This lack of community integration is also true for the subset of linked OSM stations. Yet, this subgroup records a stable growth meaning that more and more OSM stations are explicitly linked to their Wikidata counterpart. Despite the similar size of the two datasets on a global scale, they show significant discrepancies on a country level. For example, in China, Wikidata features only 39% of the stations present in OSM while having more than double the amount of stations in the United Kingdom. While the lack of stations seems reasonable considering the overall lack of stations in Wikidata, the overabundance of stations in the UK hints toward data issues, such as the misclassification of tram stops, that need more detailed analyses before integration.

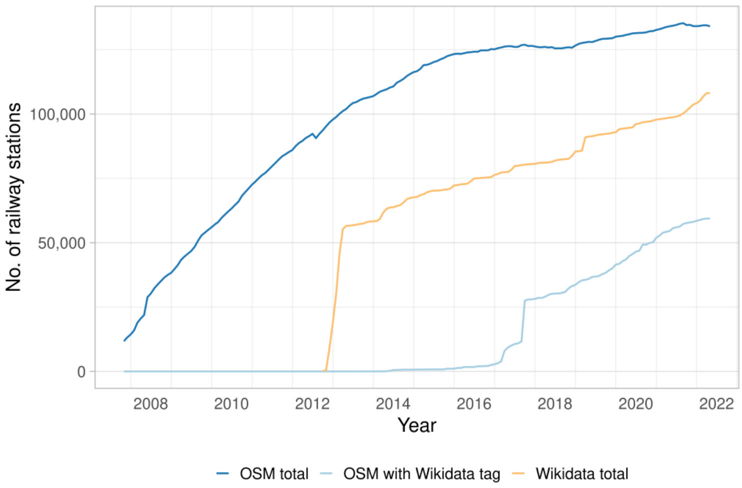


*Figure 1*. Growth Rate of OSM and Wikidata.

Regarding the properties/tags of each object, we observed that Wikidata has, on average, more properties per object than OSM. One reason could be the project's goal being knowledge collection rather than object location. Since Wikidata is a knowledge graph, it also contains links to other objects that help enrich existing objects increasing this discrepancy even further. OSM objects with links to Wikidatda have almost double the tags compared to those without links. This could be a quality indicator or an indicator that only famous stations, which are very well mapped in OSM, are also linked to Wikidata. Wikidata objects without links from OSM and geocoordinates have the least number of properties, hinting at their lower quality.

Next, we present the community analysis. There were around 1.8 million contributors in OSM in total, and 48k unique users have contributed to either creation, deletion, or updating of the railway station objects. In Wikidata, the number of overall contributors was much smaller, i.e., 24k out of which 14k have contributed to Railway objects. The revisions for Wikidata objects were around 11 times higher than that of OSM revisions. This is evident as Wikidata railway stations have more properties than OSM railway stations. This could also be because OSM contributors have a wide variety of objects to map, ranging from benches to land-use. In addition, adding a new object to the map may take priority over extensive tagging of existing objects. In contrast, Wikidata contributors may focus on details and enrichment of prominent objects of public interest. A similar trend was observed for average stations created by each contributor, wherein Wikidata contributors have created five times and, with median statistics, two times more objects than OSM contributors. This may be due to the higher number of bots and imports in Wikidata. While OSM users generally map a specific area that can only feature a limited number of railway stations, Wikidata users may import railway stations from other sources without limiting themselves to a certain geographical unit. This notion is supported when looking at the specialization of railway station contributors by calculating the share of edits made to railway stations to the total amount of edits made by a user. OSM users are less specialized than Wikidata users having only 1-2% of their edits in this domain, while Wikidata users had around 5% of their contributions in this domain. OSM users seem to often be more generalists with edits in all domains, in a certain region, while Wikidata users are more topic-driven contributors.

We notice that both communities have great potential to integrate these sources on the topic of railway stations. Although this potential increases daily with other topics reaching a mature data state in Wikidata, it is difficult to generalize our work to the entire datasets as the purpose of the datasets differs widely. OSM is focused entirely on spatial data, whereas Wikidata is a general-purpose knowledge graph. Therefore data content and style will always be tailored towards these goals and make integration a difficult task. To ensure the generalisability of our analyses to other topics, users must ensure that the data is similarly represented in both datasets in terms of the class definition and geometric representation. We also need to acknowledge that there are topics that (currently) have no potential for integration or comparison. E.g., although land-use and land-cover information is a prominent type of data in OSM, this data type is inexistent in Wikidata. The information on "what is a forest" may be present in Wikidata but is incomparable to land use polygons in OSM. To conclude, our observations of the performed analyses show that OSM can benefit from the wide range of semantic information linked to objects, while Wikidata can benefit from the precise geoinformation and completeness of OSM. The analyses can also benefit the semantic web and GIS communities by giving them insights into the datasets which can help integrate datasets. In the future, we would like to expand our work to prominent classes such as places (cities and other named locations) with additional comparison parameters.

**Acknowledgements**: This work was funded by DFG, German Research Foundation ("WorldKG", 424985896 and "IdealVGI", 424966858).

## References

[1] Tempelmeier, N., & Demidova, E. (2021). Linking OpenStreetMap with knowledge graphs—Link discovery for schema-agnostic volunteered geographic information. *Future Generation Computer Systems*, *116*, 349–364.
[2] Dsouza, A., Tempelmeier, N., & Demidova, E. (2021). Towards Neural Schema Alignment for OpenStreetMap and Knowledge Graphs. In: *Proceedings of the 20th International Semantic Web Conference, ISWC 2021*, 24-28 October 2021, Springer, Cham, 56–73.
[3] Gurtovoy, D., & Gottschalk, S. (2022). Linking Streets in OpenStreetMap to Persons in Wikidata. In: *Proceedings of The Web Conference 2022*, Lyon, France, 25-29 April 2022.
[4] Dsouza, A., Tempelmeier, N., Yu, R., Gottschalk, S., & Demidova, E. (2021). WorldKG: A World-Scale Geographic Knowledge Graph. In: *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, Queensland, Australia, 1-5 November 2021, 4475–4484.
[5] Schott, M., Herfort, B., Troilo, R., & Raifer, M. (2022). A basic guide to OSM data filtering. *GIScience News Blog.* Retrieved from http://k1z.blog.uni-heidelberg.de/2022/01/20/a-basic-guide-to-osm-data-filtering
[6] Schott, M., Grinberger, A.Y., Lautenbach, S., & Zipf, A. (2021). The Impact of Community Happenings in OpenStreetMap—Establishing a Framework for Online Community Member Activity Analyses. *ISPRS International Journal of Geo-Information*, *10*(3), 164.
[7] Schott, M., Größchen, L., & Lautenbach, S. (2022). OSM Element Vectorisation. Retrieved from https://gitlab.gistools.geog.uni-heidelberg.de/giscience/ideal-vgi/osm-element-vectorisation
[8] GIScience Research Group and HeiGIT (2022). *ohsome quality analyst*. Retrieved from https://github.com/GIScience/ohsome-quality-analyst

# Null Island - a node of contention in OpenStreetMap

Peter Mooney[1,*] and Levente Juhász[2]

[1] Department of Computer Science, Maynooth University, Maynooth, Ireland; peter.mooney@mu.ie
[2] GIS Center, Florida International University, Miami, FL, USA; ljuhasz@fiu.edu

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2022 Conference after peer-review.

Null Island refers to the location where the prime meridian meets the equator at 0 degrees longitude and 0 degrees latitude. With coordinates (0, 0), it is the origin of the WGS84 geographic coordinate system. It has been argued that Null Island can be considered a real place that is a product of our digital age [1]. Null Island's significance comes from the fact that it is erroneously associated with large amounts of geographic data that spans across geo-social media, location-based services and map databases. Even though Null Island is a fictitious, dimensionless, point object, its existence stimulates debate that elevates Null Island into a global issue worthy of serious consideration (a detailed description of associated issues is given in [1]). Members of the OpenStreetMap (OSM) project often interact with this location in various ways, and therefore understanding what Null Island means for OSM is relevant. We can find several examples of Null Island affecting OSM, such as a recent debate that arose in the talk mailing list in January 2022 with the title "Was the deletion of Null Island reasonable?" [2], where contributors argued for or against the deletion of Null Island. In addition, a web search for the term "Null Island" on the openstreetmap.org domain [3] reveals that Null Island was mentioned across the entire OSM ecosystem, including mailing lists, forums, user diaries, notes, features, changesets, wiki pages, help articles, blogs and even the Ruby on Rails codebase of the OSM website uses Null Island for testing (https://tinyurl.com/OSM-Ruby-Null). These suggest that Null Island already has a widespread reach within the OSM project.

The purpose of this study is to consider both qualitatively and quantitatively what the geographic oddity of Null Island means for OSM. No research works exist which tackle this issue in depth. Previous studies mentioning Null Island do so in a simplistic way and use the term to refer to the (0, 0) location (see e.g. [4–6]). Only a few studies recognize it as a special location and unique phenomenon [7,8] and to our knowledge, only one study tackles the issue in depth [1]. Indeed, at the outset of this work, we had expected to find many detailed treatments and inquiries into Null Island given that geographic oddities often attract research attention. In addition to contributing a robust academic study of Null Island, this work will produce a structured knowledge-based resource for the community to understand Null Island's impact on OSM.

Mooney, P., & Juhász, L. (2022). Null Island - a node of contention in OpenStreetMap.
In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.). Proceedings of the Academic Track at State of the Map 2022, Florence, Italy, 19-21 August 2022. Available at https://zenodo.org/communities/sotm-22
DOI: 10.5281/zenodo.7004488

 

Building on [1] we investigate the various ways Null Island is represented in the OSM project subsequently contributing an evidence-based narrative history on the evolution of Null Island. This includes the qualitative review of various OSM communications channels (e.g. mailing lists, discussion boards and wikis) for mentions and references to Null Island. We believe these channels help provide insights about how the OSM community contextualizes, describes and deals with Null Island. The history of special map features related to Null Island, such as node #1 (https://tinyurl.com/osm-first-node) and the node located at (0, 0) (https://tinyurl.com/OSM-Center) will also be reviewed to illustrate what actions the OSM community took in terms of adding and removing Null Island to the database. In addition to these qualitative approaches, we utilize the ohsome API [9] to extract and analyze map edits made on or near Null Island, which provides a quantitative way to assess the frequency of erroneous data added to OSM near (0, 0) as well as the semantics of such data.

Interesting patterns have already emerged from the preliminary analysis of data. The most recent mailing list debate mentioned above [2] can be summarized as follows. 17 individuals contributed 45 e-mails to the discussion between January 3 and January 10, 2022. One of the (very few) rules of OSM is that data should be verifiable, meaning that others can visit the real location of a map object and see for themselves if the data is correct. This is also known as the “ground-truth rule” [10]. Null Island as a fictional place violates this rule, therefore a popular stand in the debate is that it should not be part of OSM. This was explicitly expressed by five individuals, including a member of the authoritative Data Working Group. A counter argument is that Null Island is fundamentally similar to localities and neighborhoods, that might not exist as political or physical entities, but are known only informally to a group of people inhabiting that area. In this sense, Null Island is a place that exists in the collective consciousness of people and the name refers to the same geographic area. This justifies tagging the (0, 0) location as *place=locality* and *name=”Null Island”* in OSM. This view was explicitly supported by seven members on the mailing list. The remaining five individuals that contributed to the discussion did not take a clear stand on whether to remove or keep Null Island, but have provided arguments both for and against the deletion of it.

The full history of OSM data was extracted from the `elementsFullHistory` endpoint of the ohsome API [9] within the geographic bounding box defined by the southwest point of (-0.001, -0.001) and the northeast point of (0.001, 0.001) between January 1, 2012 and January 1, 2022. During this 10-year-long period, a feature was added, deleted or modified every three days on average within this bounding box, resulting in 1323 unique features (nodes, ways or relations). In addition, map Notes as well as GPS traces are also constantly being created, which makes Null Island and its surrounding a busy area in terms of OSM data activity.

Null Island is a socio-technological concept that has only been sparsely present in the GIScience literature so far. Our novel approach highlights how a seemingly lighthearted topic like Null Island can generate serious debates that are technological, social and even philosophical in nature. OSM and Null Island have a long tradition together with sometimes heated mapping debates resurfacing from time to time with no apparent resolution in sight. While resolving these debates is entirely in the hands of the OSM community, our research contributes to the potential resolution of them in a meaningful way by providing a factual, detailed, and accurate account of Null Island in OSM. Furthermore, while Null Island is potentially the most prominent example of a fictional place affecting maps and mapping practices, other examples also exist. For example, the most remote location on Earth, Point Nemo (which is the point in the ocean that is farthest from land) [11] is also present in OSM (https://tinyurl.com/OSM-PointNemo). Our OSM specific investigations together with a more general introduction of Null Island from both technological and social perspectives presented in [1] will help demystify the abstract concept of a fictional place that is present in real databases. Increased understanding will potentially help OSM members resolve mapping debates about “real fictional places”.

Discussion around Null Island and other fictional places is unlikely to end with this work. Indeed, our work will contribute in a technical, socio-technical and philosophical way to the Null Island story in OSM carrying with it the potential to become a catalyst for further discussions related to wider debates in OSM around mapping practices. We also believe that these discussions around Null Island span future debates and conversations about virtual geographic spaces including those in virtual open spaces such as the Metaverse. Such convergences of virtually enhanced physical and digital reality with enhanced immersive experiences will surely lead to new and exciting geographical debates around what is a real or fictional place [12].

**References**

[1] Juhász, L. & Mooney, P. (in press). I think I discovered a military base in the middle of the ocean – Null Island, the most real of fictional places. *IEEE Access*.
[2] Albrecht, E.L. (2022). Was the deletion of Null Island reasonable?. *OSM-Talk mailing list*. Retrieved from https://lists.openstreetmap.org/pipermail/talk/2022-January/087164.html
[3] Google (2022) Search query *'null island' site:openstreetmap.org*, Google search engine. Retrieved from https://www.google.com/search?q=%22null+island%22+site:openstreetmap.org
[4] Hakima, H., & Bazzocchi, M.C. (2021). Low-Thrust Trajectory Design for Controlled Deorbiting and Reentry of Space Debris. In: *2021 IEEE Aerospace Conference, 50100*, 1–10.
[5] Mound, J., Davies, C., Rost, S., & Aurnou, J. (2019). Regional stratification at the top of Earth's core due to core–mantle boundary heat flux variations. *Nature Geoscience*, *12*(7), 575–580.
[6] Hill, P.G., Stein, T.H., Roberts, A.J., Fletcher, J.K., Marsham, J.H., & Groves, J. (2020). How skilful are nowcasting satellite applications facility products for tropical Africa?. *Meteorological Applications*, *27*(6), e1966.
[7] Janowicz, K., Hu, Y., McKenzie, G., Gao, S., Regalia, B., Mai, G., Zhu, R., Adams, B., & Taylor, K. (2016). Moon landing or safari? a study of systematic errors and their causes in geographic linked data. In: *Proceedings of the Ninth International Conference on Geographic Information Science*, Montreal, Canada, 27-30 September 2016, Springer, Cham, 275–290.
[8] Kopsachilis, V., & Vaitis, M. (2021). GeoLOD: A Spatial Linked Data Catalog and Recommender. *Big Data and Cognitive Computing*, *5*(2), 17.
[9] Raifer, M, Troilo, R., Mocnik, F.-B., & Schott, M. (2021). OSHDB - OpenStreetMap History Data Analysis (version 0.7.2). Retrieved from https://doi.org/10.5281/zenodo.5512545
[10] OpenStreetMap Wiki (2022). How We Map. Retrieved from https://wiki.openstreetmap.org/wiki/How_We_Map
[11] Barnes, R. (2020) Optimal orientations of discrete global grids and the Poles of Inaccessibility. *International Journal of Digital Earth 13(7),* 803–816.
[12] Gordon, E. (2008). The Geography of Virtual Worlds: An Introduction. *Space and Culture*, *11*(3), 200–203.

# Increasing OSM Data Accessibility with the Analysis-Ready Daylight Distribution of OSM: Demonstration of Cloud-Based Assessments of Global Building Completeness

Jennings Anderson[1,*] and Timera W. Omidire[1]

[1] Meta Platforms, Inc., USA; jenningsa@fb.com, omidire@fb.com

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2022 Conference after peer-review.

Despite being one of the most open and freely available spatial datasets, OpenStreetMap (OSM) data accessibility remains a challenge. Data accessibility measures how easily end-users can access and use a given dataset for their needs [1]. With its unique data structure of nodes, ways, and relations connected via tags, OSM data is not immediately consumable by standard geospatial analysis tools. OSM Pre-analysis workflows require OSM data to be downloaded, parsed, and converted into more common formats, which means that novice end-users of OSM may lack the experience to readily access and use OSM data in decision-making.

Incorporating communities into spatial decision-making processes, such as mapping, is important because a). community members are experts on their communities and b). have a larger stake in final solutions which directly impacts their lived-experiences [2]. OSM empowers a variety of communities, including local governments [3], digital humanitarian groups [4], and even student groups [5], to help navigate and understand places of respective importance.

Research by Nirandjan et al. recently lowered barriers to using OSM data as a reference dataset of critical infrastructure [6]. After categorizing and quantifying particular types of OSM features, the authors released the data in formats more common in geospatial analysis, such as GeoTiffs [6]. This article's popularity (ranked 90th percentile on the publisher's website) demonstrates the importance of making OSM data—and datasets derived from OSM—more accessible by means of familiar data structures compatible with common tools. If OSM data were more accessible for analysis, would we see it used in more geospatial research and innovation at large [7]?

While many community-maintained tools exist to convert, extract, and download OSM data, each requires domain knowledge of the unique OSM data structure (nodes, ways, and relations). Furthermore, working at the country or planet-scale requires extensive computational resources. To further lower the barriers to entry for OSM data analysis and

Anderson, J., & Omidire, T.W. (2022). Increasing OpenStreetMap Data Accessibility with the Analysis-Ready Daylight Distribution of OpenStreetMap: A Demonstration of Cloud-Based Assessments of Global Building Completeness.
In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.). Proceedings of the Academic Track at State of the Map 2022, Florence, Italy, 19-21 August 2022. Available at https://zenodo.org/communities/sotm-22
DOI: 10.5281/zenodo.7004501
 

extraction, we created the Analysis-Ready Daylight OpenStreetMap Distribution (ARD-OSM). Freely available to anyone, it is published on the registry of open data (RODA) on Amazon Web Services (AWS) [8]. Containing 1B OSM features, ARD-OSM is optimized for use with Amazon Athena, a serverless interactive query engine on AWS. Additionally, ARD-OSM has resolved the OSM data format into common geometries such as points, lines, and polygons. The dataset also includes pre-computed valuable attributes such as length, surface area, quadkeys, and geographic bounding boxes which are stored as additional metadata. To demonstrate the analytical capabilities of this dataset, next, we will perform a global OSM building density assessment.

Building density is a common metric in OSM quality research, often used to assess map coverage and completeness, such as studied by Yeboah et al. [9]. Measuring building density requires counting all of the buildings within a defined unit of spatial analysis. We use zoom-level 11 map tiles to create an analysis grid that encompasses the global built environment in fewer than 1M tiles. Then, we divide the building count by the area of each map tile to obtain the number of buildings mapped per square kilometer.

Since every feature in ARD-OSM includes the zoom-level 15 quadkey of the map tile in which it exists, we can use a more efficient SQL GROUP BY expression instead of a geospatial operator for aggregation. After setting up the *analysis_ready_daylight* table in Amazon Athena, this is the query used to count the number of buildings in each zoom-level 11 map tile:

```
SELECT      substr(quadkey, 1, 11) as z11_tile,
            count(id) as number_of_buildings
FROM  analysis_ready_daylight
WHERE tags['building'] IS NOT NULL AND release = 'v1.12'
GROUP BY substr(quadkey, 1, 11)
```

In May 2022, running in AWS region us-east-1, this query took 15 seconds and cost USD $0.10. In this way, a global accounting of all the buildings in OSM can be performed quickly, cost-effectively, and by anyone with a basic knowledge of common SQL. The results of this query show the density of mapped buildings in OSM to be highest in Europe with additional areas of high density where Humanitarian mapping campaigns have been active such as Nepal, South Eastern Asia, and isolated parts of Africa. This is consistent with the findings of Herfort et al. [10].

How should these densities be interpreted? Do denser regions have higher completeness in which most or all buildings are mapped? Building density is an intrinsic data quality measure, to further contextualize these findings, we need to perform an extrinsic assessment by comparing our results against an external dataset. A recent study confirmed the viability of referencing population data for building density assessment [11]; and Orden et al. demonstrate a three-step methodology using Meta's High Resolution Settlement Layer (HRSL) first requiring both vectorization and spatial aggregation to assess building completeness with respect to population in both the Philippines and Madagascar [12].

Because the HRSL is also published via RODA [13], it can be easily *joined* to our results. Once HRSL data is incorporated to obtain a measure of *buildings mapped per square kilometer per person*, we find that parts of Europe remain in the top tiers of density with the most buildings mapped per person. Nepal and many parts of South Eastern Asia, however, are no longer in the same top tier of map coverage. Figure 1 depicts these differences. While there are many mapped buildings, the higher populations of these regions reveals that there are still many areas where the buildings have yet to be mapped. This yields a generally lower level of completeness than initially identified, which remains consistent with the findings of Herfort et al. [10]. Additionally, parts of the United States and New Zealand actually appear more complete with areas of lower density coinciding with regions of lower population, yielding a higher measure of map completeness than before.

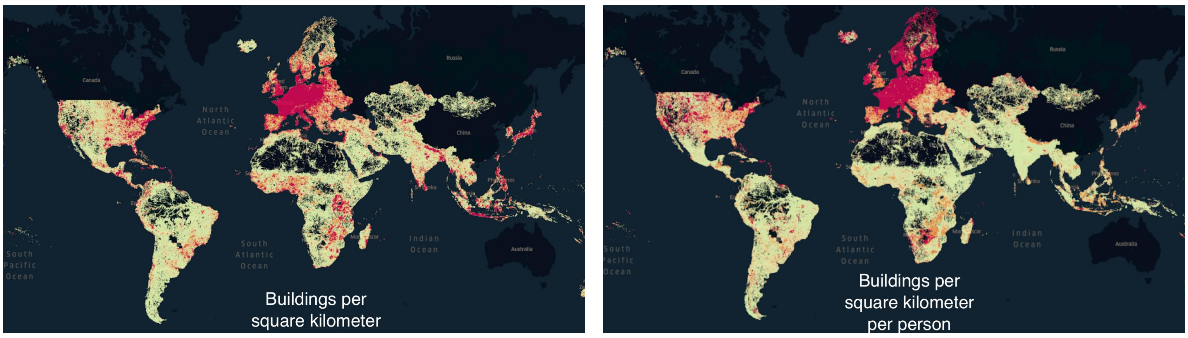


*Figure 1*. Subtle differences in building density in OSM when incorporating population density. Areas in red have highest density while areas in yellow have lower densities.

This case study cheaply and easily reproduced popular methods for both intrinsic and extrinsic data quality assessments of OSM building coverage without needing to download nor pre-process any OSM data. The analysis was done completely in the cloud on AWS using free and open data in RODA. Additional metadata in ARD-OSM enabled the query to run efficiently and cost-effectively. The same methodology can be applied to investigations of any other object type in OSM from hospitals to ice cream shops. We also recognize that ARD-OSM does not solve the needs of researchers looking to work with OSM history data. Other tools such as Ohsome, which utilizes the OpenStreetMap History Database are better suited for those types of historical analyses [14]. Researchers competent in the Java programming language can leverage the Ohsome API to interrogate the complete history of OSM features. Such investigations go beyond the current scope of ARD-OSM, which is optimized for efficient planet-scale exploration of the latest version of the map. Future innovations to the dataset and process could include pre-computing the complete history of each feature, but for the time-being, we recommend using Ohsome for such investigations.

Additionally, while ARD-OSM is currently available on AWS services, the raw data is a series of Parquet files intended to be used with PrestoDB, therefore, users are not locked into AWS. We see the creation of ARD-OSM as an example of what is possible when OSM researchers utilize distributed cloud-based database technologies and hope that future researchers can expand on these findings.

ARD-OSM contains 1B features—nearly every OpenStreetMap object—in common geospatial feature types such as points, lines, and polygons. To additionally aid researchers, features are enriched with additional metadata describing their location and physical attributes such as length or surface area. From a data accessibility perspective, we anticipate ARD-OSM in future research and innovation, curated by a wide range of end-users, to readily integrate OSM data for decision-making processes which bring communities closer together.

## References

[1] Open Data Charter (2015). Principles. Open Data Charter. Retrieved from https://opendatacharter.net/principles
[2] Keenan, P.B., & Jankowski, P. (2018). Spatial decision support systems: Three decades on. *Decision Support Systems, 116*(2019), 64–76.
[3] Johnson, P.A. (2017). Models of direct editing of government spatial data: challenges and constraints to the acceptance of contributed data. *Cartography and Geographic Information Science, 44*,(2), 128–138.
[4] Palen, L., Soden, R., Anderson, T.J., & Barrenechea, M. (2015). Success & Scale in a Data-Producing Organization: The Socio-Technical Evolution of OpenStreetMap in Response to Humanitarian Events. In: *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, Seoul, Republic of Korea, 18-23 April 2015, 4113–4122.
[5] Maidin, M.A.A., Ahmad, F., Abidin, N.I., Suhaili, J., Awang, M., Rahman, M.A.A., Musa, M.K., Hamidon, N., Yusop, F.M., Syazwan, M.M.S., Harun, H., Hamid, N.H.A., & Kamil, N.A. (2021). Design campus map using OpenStreetMap digital software. In: Zaini, M.A.A., Jusoh, M., & Othman, N. (Eds.) *Proceedings of the 3rd International Conference on Separation Technology*, Springer, Singapore, 113–129.
[6] Nirandjan, S., Koks, E.E., Ward, P.J., & Aerts, J.C.J.H. (2022). A spatially-explicit harmonized global dataset of critical infrastructure. *Scientific Data 9*(1), 11–13.
[7] National Academy of Sciences (US), National Academy of Engineering (US) and Institute of Medicine (US) Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age (2009) *Ensuring the integrity, accessibility, and stewardship of research data in the digital age*. Washington, DC, National Academies Press.
[8] Daylight Map Distribution of OpenStreetMap. Retrieved from https://registry.opendata.aws/daylight-osm
[9] Yeboah, G., Porto de Albuquerque, J., Troilo, R., Tregonning, G., Perera, S., Ahmed, S.A.K.S., Ajisola, M., Alam, O., Aujla, N., Azam, S.I., Azeem, K., Bakibinga, P., Chen, Y.-F., Choudhury, N.N., Diggle, P.J., Fayehun, O., Gill, P., Griffiths, F., Harris, B., Iqbal, R., Kabaria, C., Ziraba, A.K., Khan, A.Z., Kibe, P., Kisia, L., Kyobutungi, C., Lilford, R.J., Madan, J.J., Mbaya, N., Mberu, B., Mohamed, S.F., Muir, H., Nazish, A., Njeri, A., Odubanjo, O., Omigbodun, A., Osuh, M.E., Owoaje, E., Oyebode, O., Pitidis, V., Rahman, O., Rizvi, N., Sartori, J., Smith, S., Taiwo, O.J., Ulbrich, P., Uthman, O.A., Watson, S.I., Wilson, R., & Yusuf, R. (2021). Analysis of OpenStreetMap Data Quality at Different Stages of a Participatory Mapping Process: Evidence from Slums in Africa and Asia. *ISPRS International Journal of Geo-Information, 10*(4), 265.
[10] Herfort, B., Lautenbach, S., Porto de Albuquerque, J., Anderson, J., & Zipf, A. (2021). The evolution of humanitarian mapping within the OpenStreetMap community. *Scientific Reports, 11*, 3037.
[11] Zhang, Z., Zhou, Q., Brovelli, M.A., & Wanjing, Li (2022). Assessing OSM building completeness using population data. *International Journal of Geographical Information Science*, *36*(7), 1443–146.
[12] Orden, A., Flores, R.A., Faustino, P., & Samson, M.S. (2020). Measuring OpenStreetMap building footprint completeness using human settlement layers. In: Minghini, M., Coetzee, S., Juhász, L., Yeboah, G., Mooney, P., & Grinberger, A.Y. (Eds.) *Proceedings of the Academic Track at the State of the Map 2020 Online Conference*, 4-5 July 2020.
[13] High Resolution Population Density Maps + Demographic Estimates by CIESIN and Meta. Retrieved from https://registry.opendata.aws/dataforgood-fb-hrsl.
[14] Raifer, M., Troilo, R., Kowatsch, F., Auer, M., Loos, L., Marx, S., Przybill, K., Fendrich, S., Mocnik, F.B., & Zipf, A. (2019). OSHDB: a framework for spatio-temporal analysis of OpenStreetMap history data. *Open Geospatial Data, Software and Standards*, *4*(1), 1–12.

# OSM for sustainable transport planning

Greta Timaite[1]*, James Hulse[1] and Robin Lovelace[1,2]

[1] Leeds Institute for Data Analytics, University of Leeds, Leeds, United Kingdom; g.timaite@leeds.ac.uk, j.j.c.hulse@leeds.ac.uk
[2] Institute for Transport Studies, University of Leeds, Leeds, United Kingdom; r.lovelace@leeds.ac.uk

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2022 Conference after peer-review.

One of the key domains in which OpenStreetMap (OSM) data has been utilised is transport planning [1]. OSM has been used in agent-based transport simulation [2] and routing [3], including cycling [4], walking [5], wheeling [6], and blind pedestrian routing [7]. Another application of OSM data is to support evidence-based investments in sustainable transport infrastructure. In a recent (2021) paper Nelson et al. [8] argue that OSM has the potential to become a primary source of data on infrastructure across the globe.

Regardless of OSM's potential to become a primary source of data on transport infrastructure in general, its potential in active travel infrastructure planning is yet to be realised. One of the reasons behind this lag might be linked to the perceived unreliability of open-access crowdsourced data [9] and the fact that planning for active travel requires high resolution and rich data, e.g. sidewalk widths, geometries and surface roughness data. The quality of OSM has received extensive examination [1] in which the question concerning data completeness plays a significant role because, it is argued, the mappers are not coordinated to guarantee systematic coverage [10]. To address this issue, Barrington-Leigh and Millard-Ball [11] assessed OSM road completeness and found that globally over 80% of roads are mapped. Problematically, their assessment focused on roads designed for motor traffic, thus excluding other modes of transport. This gap has been partially addressed by Ferster et al. [12] who examined OSM cycling infrastructure in Canada. They have not, however, considered the infrastructure from the accessibility perspective. Moreover, there seems to exist no equivalent study using OSM data for pedestrian infrastructure planning.

Yet, open-access crowdsourced data, such as OSM, can support increasing demand for local evidence to inform transport policies. This is important in the context of countries such as the UK, which has bold walking and cycling targets: 50% of trips made by walking and cycling in towns and cities by 2030 [13]. Such targets require a shift in transport planning, away from provision for motorised modes towards more sustainable active modes of travel, such as walking, wheeling, and cycling [14]. The importance of localising interventions to meet the needs of local communities has been outlined in both policy [16] and academic papers [17]. An increase in citizen engagement in decision-making could be achieved through the encouragement of "produsage" – a model in which citizens both produce and use data [18].

Timaite, G., Hulse, J., & Lovelace, R. (2022). OSM for sustainable transport planning.
In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.). Proceedings of the Academic Track at State of the Map 2022, Florence, Italy, 19-21 August 2022. Available at https://zenodo.org/communities/sotm-22
DOI: 10.5281/zenodo.7004509
 

Acknowledging the potential of OSM to boost citizen participation, the OpenInfra project aims to address the gap concerning the potential of OSM in transport research. The project started by examining the existing OSM tags relevant to active travel infrastructure in England with a focus on West Yorkshire, Greater Manchester, Greater London, and Merseyside. The data was imported using osmextract [19], an R package, and explored using a range of exploratory data analysis (EDA) techniques. Reproducible code that generates all the figures presented here can be found on GitHub: https://github.com/udsleeds/openinfra/tree/main/sotm2022.

Given the extensive use of OSM data in transport research, it is not surprising that OSM provides a comprehensive active travel network, yet there is a lack of specification concerning the type of infrastructure that is present. For instance, cycleways and footways constitute about 1/3 of all the mapped highways on which one can legally walk, wheel or cycle but only a few percent of the cycleways and footways have tags detailing their type. The data gets even scarcer in the context of accessible infrastructure planning. For example, there is a lot of missing information on the presence and type of kerbs – a street element that might make the movement of a wheelchair user more challenging [20].

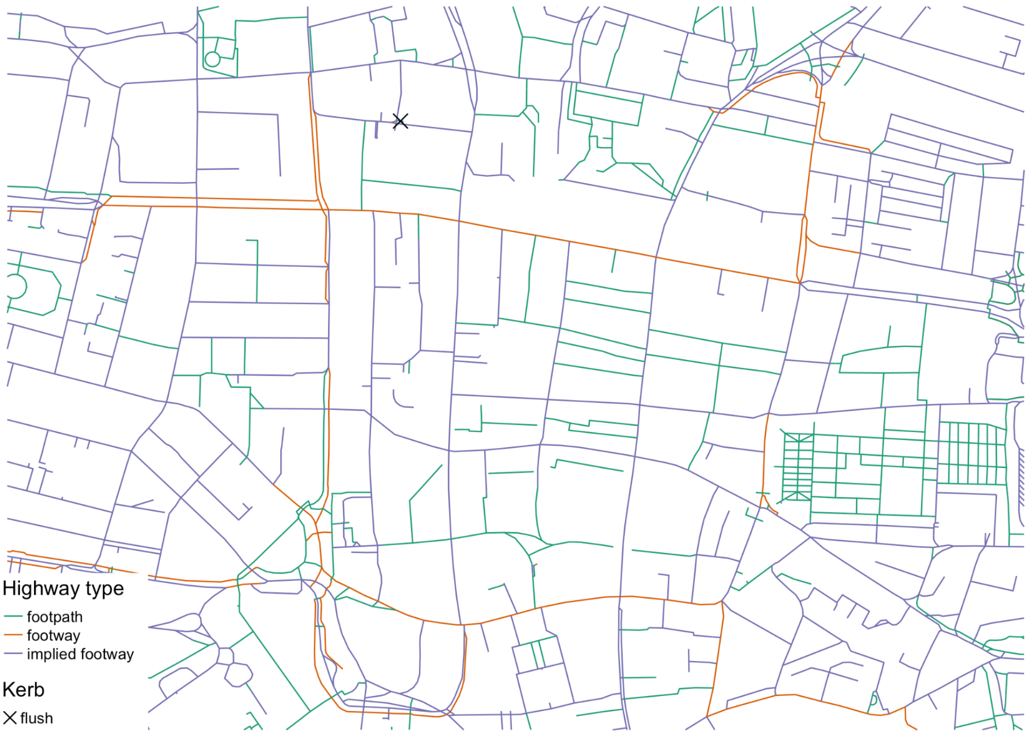


*Figure 1.* Footpaths, footways, implied footways, and kerbs in central Leeds as defined by the Inclusive Mobility guide.

The missing data currently limits the use of OSM data in active travel planning, however this does not mean that the use of OSM data should be dismissed. Following Nelson et al.'s [8] argument that it is important to make crowdsourced data more actionable, we decided to recategorise OSM data based on the Inclusive Mobility (IM) [16], a guide outlining the best practices for creating inclusive pedestrian infrastructure in the UK. For this, a function has been written (documentation: https://udsleeds.github.io/openinfra/articles/im_get.html). However, the function provides a simplification of the IM guide for a couple of reasons. The first one could be considered in terms of definitional discrepancies. For instance, the guide defines footways as “pavements adjacent to roads”, but this is not easily extracted from the OSM in which *highway=footway* is a generic tag and often there is no further refinement (e.g., *sidewalk=**) to determine if it is a pavement adjacent to a road. Another reason is linked to assigned values. For example, the guide identifies six tactile paving surfaces but OSM focuses on the presence/absence of tactile paving, thus limiting how much information can be extracted from the data.

One potential application of the IM function could be to explore the existence and geographic distribution of accessibility indicators, such as the presence of footways, footpaths, or kerbs (see Figure 1). Yet, more interesting results can be produced by using recategorised OSM data in conjunction with other datasets that would improve the understanding of street accessibility. To illustrate this, an open-access Leeds Central Council Footfall data was used [21]. We reasoned that the locations at which footfall data was collected are heavily used by pedestrians, thus demonstrating the need to ensure inclusivity of spaces. 5 unique streets were identified, which resulted in 35 linestrings in OSM. Then, a basic index of accessibility, defined by 5 indicators, was created. If a street meets all 5 requirements, it receives a maximum index of 5; for every missing indicator the score is dropped by 1. For example, if a linestring classifies as a footway, footpath, or implied footway based on the IM guide then it receives +1. If it also has an even surface, it receives an additional +1, thus totaling to +2. If none of the other conditions are met then the final index score is 2. Following this, 19 out of 35 mapped linestrings scored 2 while the rest had an index of 1. This example does not necessarily show that the streets are inaccessible because the missing data makes it hard to make a fair judgement. However, we would argue that this is a space for OSM to produce more readily actionable insights regarding transport infrastructure, especially if joined with other (open) datasets that would help to overcome some of its current data limitations.

The following steps of the OpenInfra project are focused on scaling up. The goal is to produce ‘OSM transport infrastructure data packs’ for transport authorities in the UK to support the uptake of open-access data, such as OSM, in transport planning. We believe that our project to make OSM data more ‘analysis ready’ for transport planners will make the process of transport planning more transparent, reproducible, and participatory. Increased use of OSM data by practitioners could lead to a positive feedback loop in which more people contribute to OSM, further raising the profile. Indeed, with sustained financial and political support, the uptake of and social investment in crowdsourced data could lead to more evidence-based and therefore more effective investment in sustainable transport infrastructure and, ultimately, more active travel leading to health and environmental benefits [13]. OSM specifically has the potential to provide localised insights on the existing transport infrastructure and facilitate more inclusive and accessible transport planning.

**Acknowledgements**: This work was supported by the ESRC funded Consumer Data Research Centre (CDRC) under grant references ES/S007164/1 and ES/L011891/1.

## References

[1] Jokar Arsanjani, J., Zipf, A., Mooney, P., & Helbich, M. (2015). An Introduction to OpenStreetMap in Geographic Information Science: Experiences, Research, and Applications. In: Jokar Arsanjani, J., Zipf,

A., Mooney, P., & Helbich, M. (Eds.) *OpenStreetMap in GIScience: Experiences, Research, and Applications*, Springer, Cham, 1–15.
[2] Ziemke, D., Kaddoura, I., & Nagel, K. (2019). The MATSim Open Berlin Scenario: A multimodal agent-based transport simulation scenario based on synthetic demand modeling and open data. *Procedia Computer Science, 151*, 870–877.
[3] Luxen, D., & Vetter, C. (2011). Real-time routing with OpenStreetMap data. In: *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 513–516.
[4] Yeboah, G., & Alvanides, S. (2015). Route Choice Analysis of Urban Cycling Behaviors Using OpenStreetMap: Evidence from a British Urban Environment. In: Jokar Arsanjani, J., Zipf, A., Mooney, P., & Helbich, M. (Eds.) *OpenStreetMap in GIScience: Experiences, Research, and Applications*, Springer, Cham, 189–210.
[5] Zielstra, D., & Hochmair, H.H. (2012). Using Free and Proprietary Data to Compare Shortest-Path Lengths for Effective Pedestrian Routing in Street Networks. *Transportation Research Record, 2299*(1), 41–47.
[6] Neis, P. (2015). Measuring the Reliability of Wheelchair User Route Planning based on Volunteered Geographic Information. *Transactions in GIS, 19*(2), 188–201.
[7] Cohen, A., & Dalyot, S. (2021). Route planning for blind pedestrians using OpenStreetMap. *Environment and Planning B: Urban Analytics and City Science, 48*(6), 1511–1526.
[8] Nelson, T., Ferster, C., Laberee, K., Fuller, D., & Winters, M. (2021). Crowdsourced data for bicycling research and practice. *Transport Reviews, 41*(1), 97–114.
[9] Basiri, A., Haklay, M., Foody, G., & Mooney, P. (2019). Crowdsourced geospatial data quality: Challenges and future directions. *International Journal of Geographical Information Science, 33*(8), 1588–1593.
[10] Haklay, M. (2010). How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets. *Environment and Planning B: Planning and Design, 37*(4), 682–703.
[11] Barrington-Leigh, C., & Millard-Ball, A. (2017). The world's user-generated road map is more than 80% complete. *PLOS ONE, 12*(8), e0180698.
[12] Ferster, C., Fischer, J., Manaugh, K., Nelson, T., & Winters, M. (2020). Using OpenStreetMap to inventory bicycle infrastructure: A comparison with open data from cities. *International Journal of Sustainable Transportation, 14*(1), 64–73.
[13] UK Department for Transport (2020). *Gear change: A bold vision for cycling and walking*.
[14] UK Department for Business, Energy & Industrial Strategy (2021). *Net Zero Strategy: Build Back Greener*.
[15] UK Department for Transport (2020). *Cycle infrastructure design (LTN 1/20)*.
[16] UK Department for Transport (2021). *Inclusive mobility: A guide to best practice on access to pedestrian and transport infrastructure*.
[17] Aldred, R., Woodcock, J., & Goodman, A. (2016). Does More Cycling Mean More Diversity in Cycling? *Transport Reviews, 36*(1), 28–44.
[18] Boularouk, S., Josselin, D., & Altman, E. (2017). Ontology for a voice transcription of OpenStreetMap data: The case of space apprehension by visually impaired persons. *World Academy of Science, Engineering and Technology*, London, UK, hal-01533064.
[19] rOpenSci Project (2021). *osmextract*. Retrieved from https://github.com/ropensci/osmextract
[20] Timaite, G., Lovelace, R., & Houlden, V. (2022). The potential of OpenStreetMap for (accessible) active travel planning. *30th Annual Geographical Information Science Research UK (GISRUK)*, 5-8 April 2022. Retrieved from https://doi.org/10.5281/zenodo.6408417
[21] Leeds City Council (2021). Leeds City Council Footfall Camera Aggregated Data. Retrieved from https://data.cdrc.ac.uk/dataset/leeds-city-council-footfall-camera-aggregated-data

# OSM Sidewalkreator - A QGIS plugin for automated sidewalk drawing for OSM

Kauê Vestena[1,*], Silvana Camboim[1] and Daniel Santos[2]

[1] Department of Geomatics, Federal University of Paraná, Curitiba, Brazil; kauemv2@gmail.com, silvanacamboim@gmail.com
[2] Cartographic Engineering Sector, Military Engineering Institute, Rio de Janeiro, Brazil; daniel.rodrigues@ime.eb.br

Sidewalks are a relevant part of the living space in urban environments. The existence of sidewalks and their condition is fundamental to locomotion in general and is critically important in mobility groups such as cyclists, wheelchair users, blind people, the elderly, and children. Also, the displacement along sidewalks can ensure safety from traffic, contributing to the well-being of citizens [1].

There is still an open debate about the best way to represent sidewalks in the OpenStreetMap community. Some users claim that they should be represented only as tags of the streets, using compound tags such as "*sidewalk:left/right:surface=**", arguing that over-representation can pollute the map and create unnecessary complexity [2]. On the other hand, Biagi [3] and many OSM users nowadays [4] have been showing the representation of sidewalks as separated geometries as allegedly their best representation in OSM. There are many listed advantages [4]: crossings may be represented as lines, with the kerb interfaces as points (8 in a regular 4-way intersection); the actual traversing length will be represented appropriately (as it will include block corners and crossings); independence from the digitizing direction, as "left" and "right" may swap if someone inverts the way direction; ease of representation for pedestrian islands. Moreover, some cases cannot be represented correctly using the tag scheme or will need some cumbersome solutions, as it shall represent the portion of the sidewalk orthogonally projected from the street. Therefore, if a property differs on both sides, one may need to split the highway into four segments to represent it correctly. There are also other issues, e.g. geometric properties such as the distance from the sidewalk to the street will stay unclear.

Regardless of the form chosen for representation, there are still few mapped sidewalks. For example, according to Taginfo [5], as of April of 2022, there are approximately 201 million ways with the *highway* key, but only 16.8 million (7.61%) are tagged as *highway=footway*, considering the key *footway*, there are only 4.8 million (2.45% of 201 million) ways tagged as such (58% *sidewalk*, 41% *crossing*), considering the tag *sidewalk=**, there are only 2.6 million (1.3% of 201 million) ways tagged. So, considering that most features are located in urban environments [6,7], where the major part of streets may have a sidewalk, the sidewalks are underrepresented in both schemes. Historically, it has been an issue, as [8] showed that in Berlin, only 5.6% of the Highways have a *sidewalk* tag, growing to only 8.2% in 2017 [9].

Vestena, K., Camboim, S., & Santos, D. (2022). OSM Sidewalkreator - A QGIS plugin for automated sidewalk drawing for OSM.
In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.). Proceedings of the Academic Track at State of the Map 2022, Florence, Italy, 19-21 August 2022. Available at https://zenodo.org/communities/sotm-22
DOI: 10.5281/zenodo.7004523

 

Recently, the mapping of sidewalks has grown in importance among the OSM and academic [10–13] communities. For example, the Open Sidewalks Initiative [14]. They are both a community and a project, providing dedicated mapping with an elaborate scheme on how to map in a pedestrian-centred way, but only manually. Unfortunately, drawing sidewalks and crossings is time-consuming and can be error-prone. This effort can be inferred by examining the OpenSidewalk's Tasking Manager [15]. In the most near-completion project [16], each task has taken an average of 9.4 minutes to be mapped, but 22.7 minutes to be validated. Thus, considering just crossing mapping, for the 1046 existing tasks, it will take approximately 163.9 hours of mapping and 395.7 hours of validation. This total encompasses an area of just approximately 6.17 $km^2$, only 0.65% of the urban area of the city of Sao Paulo, for example.

To cover this gap, we propose a Github-hosted, fully open-source QGIS Plugin entitled "OSM SidewalKreator" [17], which aims to automatically draw for OSM the geometries of sidewalks, crossings and kerb crossing interfaces, along with the basic descriptive tags. This tool allows the user to control and supervise the entire process. It contains a user-friendly GUI (Guided User Interface, presented in Figure 1) that enables and disables the buttons according to the step in the process. The plugin methodology, encompassing the steps that the plugin runs through are basically: 1) Fetch Interest OSM data (highways and optionally buildings and addresses, when available) from a bounding polygon given by the user; 2) Generate a table for standard widths for the values for the *highway=** tag, to provide widths to highways that have no *width=** tag; 3) Remove ways that aren't streets, according to a value of width below 0.5 m in the table; and split into segments at road intersections; 4) Generate the sidewalk geometries based on a per-segment buffer (optionally checking if it doesn't overlaps buildings), dissolve, and a negative then positive buffer to create rounded corners (if wanted by the user) and then finally extract line geometries (inner holes outlines); 6) Generate the crossings geometry, using a vector that that grows iteratively in a direction perpendicular to the segment or parallel to adjacent segments (according to user's choice), until an intersection with the sidewalks (dissolved as a single multipart geometry) is found, there are also filtering options, to avoid badly generated crossings, such as too long geometries or too close to other crossings; 7) Split the sidewalk geometries, since sidewalks properties (smoothness, surface material, etc.) may differ many times in the same block, it can split based on voronoi polygons of building centroids and/or addresses, or with a minimal length, minimal number of segments, only block façades or don't split; 8) Output all features to a single .geojson ready to be imported at JOSM editor, where more inspection on generated data can be carried out, e.g. some manual editing. The plugin also outputs other auxiliary files as a .txt with a pre-filled changeset comment and intermediary files for debugging. The tool's first use test was performed in April 2022 to map the Campus Centro Politécnico of the Federal University of Paraná, with transportation engineering students, for the Horus Nav Project [18], an open-source tool for route optimization for blind people.

The key idea that guides the present work is to provide a tool to help intermediary to advanced users to speed up the tedious task of manually drawing sidewalks, crossings and kerbs. The tool does not intend to be a fully automated (a challenging task [12,13]) one-click solution but always calls the user to check out the results, step-by-step, using the resourcefulness from QGIS and JOSM. After the data is imported, the user is encouraged to carry out a validation project on a Tasking Manager. The present work also advocates for the representation of sidewalks as separate geometries as the best available way to represent the sidewalk network. However, the tag scheme still can be helpful in some specific situations, such as explicitly representing that some highway does not have a sidewalk on one or both sides. Taking advantage of previously available data, including tag schema, is one of the features that might be included in future releases of OSM SidewalKreator and integrating properly with previously drawn sidewalks and crossing restrictions. Finally, by joining awareness in mapping communities, detailed instructions, and tools for automation, we can deepen, improve, and increase the amount of sidewalk presence in OSM towards a more pedestrian-centred way of mapping, providing support to applications like optimized routing for pedestrians, with dedicated profiles for people with disabilities. In the long term, the results of this project are aimed at building a solid foundation for improving accessibility and mobility in cities around the world.

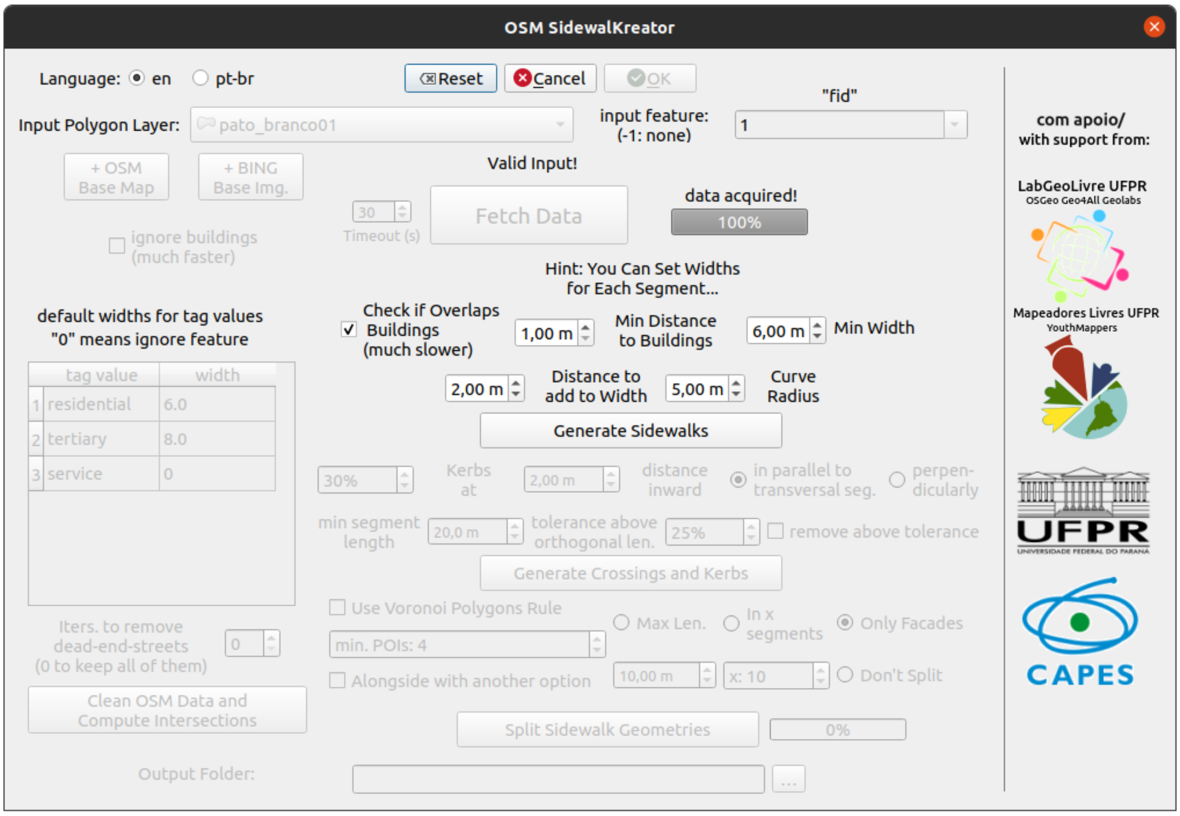


*Figure 1*. OSM SidewalKreator Current GUI.

## References

[1] US Department of Transportation, Federal Highway Administration - FHWA Safety Program (2013). Safety benefits of walkways, sidewalks, and paved shoulders. Retrieved from https://Safety.Fhwa.Dot.Gov/Ped_bike/Tools_solve/Walkways_trifold

[2] OpenStreetMap Wiki (2011). Proposed features/Sidewalk as separate way. Retrieved from https://wiki.openstreetmap.org/w/index.php?oldid=2312766

[3] Biagi, L., Brovelli, M.A., & Stucchi, L. (2020). Mapping The Accessibility In Openstreetmap: A Comparison Of Different Techniques. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B4*, 229–236.

[4] OpenStreetMap Wiki. (2016). Proposed features/sidewalk schema. OpenStreetMap Wiki. Retrieved April from https://wiki.openstreetmap.org/wiki/Proposed_features/sidewalk_schema

[5] OpenStreetMap contributors (2022). Taginfo. Retrieved from https://taginfo.openstreetmap.org

[6] Neis, P., & Zielstra, D. (2014). Recent developments and future trends in volunteered geographic information research: The case of OpenStreetMap. *Future Internet, 6*(1), 76–106.
[7] Camboim, S., Bravo, J., & Sluter, C. (2015). An investigation into the completeness of, and the updates to, OpenStreetMap data in a heterogeneous area in brazil. *ISPRS International Journal of Geo-Information, 4*(3), 1366–1388.
[8] Mobasheri, A., Bakillah, M., Rousell, A., Hahmann, S., & Zipf, A. (2015). On the completeness of sidewalk information in OpenStreetMap, a case study of Germany. In: *Proceedings of the 18th AGILE International Conference on Geographic Information Science*, Lisbon, Portugal, 9-12 June 2015.
[9] Mobasheri, A., Sun, Y., Loos, L., & Ali, A. (2017). Are crowdsourced datasets suitable for specialized routing services? Case study of OpenStreetMap for routing of people with limited mobility. *Sustainability, 9*(6), 997.
[10] Ertz, O., Fischer, A., Ghorbel, H., Hüsser, O., Sandoz, R., & Scius-Bertrand, A. (2021). Citizen Participation & Digital Tools To Improve Pedestrian Mobility In Cities. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVI-4/W1*, 29–34.
[11] Jiang, Y., Lobo, M. J., Christophe, S., & Jouffrais, C. (2021). Mapping road crossings for visually impaired people. *30th International Cartographic Conference (ICC 2021)*, Florence, Italy, 14-18 December 2021, 1–2.
[12] Gjeruldsen, E. (2020). *Detecting sidewalks in OpenStreetMap using machine learning LATEX Version*. Master's thesis, Østfold University. Retrieved from https://hdl.handle.net/11250/2676606
[13] Mobasheri, A., Huang, H., Degrossi, L., & Zipf, A. (2018). Enrichment of OpenStreetMap data completeness with sidewalk geometries using data mining techniques. *Sensors, 18*(2), 509.
[14] Taskar Center for Accessible Technology, University of Washington (2021). OpenSidewalks. Retrieved from https://www.opensidewalks.com
[15] Taskar Center for Accessible Technology & Humanitariam OpenStreetMap Team (2021). OpenSidewalks Tasking Manager. Retrieved from https://tasks.opensidewalks.com
[16] Taskar Center for Accessible Technology & Humanitariam Openstreetmap Team (2021). OpenSidewalks Mapping statistics Vila Mariana, Sao Paulo, Brazil - Crossings. Retrieved from https://tasks.opensidewalks.com/projects/51/stats
[17] Vestena, K.M. (2021). OSM SidewalKreator (version 0.9). Retrieved from https://github.com/kauevestena/osm_sidewalkreator
[18] Klavier, D. (2022). Horus Nav (version 0.1). Retrieved from https://github.com/klaiiver/HorusNav

# Inequalities in the completeness of OpenStreetMap buildings in urban centers

Benjamin Herfort[1,2]*, Sven Lautenbach[1], João Porto de Albuquerque[3], Jennings Anderson[4] and Alexander Zipf[2]

[1] Heidelberg Institute for Geoinformation Technology, Heidelberg, Germany; benjamin.herfort@heigit.org, sven.lautenbach@heigit.org
[2] GIScience Chair, Institute of Geography, Heidelberg University, Heidelberg, Germany; zipf@uni-heidelberg.de
[3] Urban Big Data Centre, University of Glasgow, United Kingdom; Joao.Porto@glasgow.ac.uk
[3] Meta Platforms Inc., United States; jennings.anderson@gmail.com

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2022 Conference after peer-review.

The collaborative maps of OpenStreetMap (OSM) have become a major source of geospatial baseline data for humanitarian organisations, companies and public authorities. Describing the elements of spatial data quality (e.g. positional accuracy, completeness, temporal quality) for the OSM dataset is a key prerequisite to provide the potential stakeholders with the necessary information to decide on the fitness for use of a data set for their particular application [1]. Better spatial data quality assessment would promote the adoption and (right) usage of new sources of data such as OSM and data products based on OSM. A large community of researchers has analyzed the quality of OSM data in comparison to authoritative reference data sets, by means of remote sensing and using intrinsic measures [2–4]. It has been acknowledged that the OSM data in general is strongly biased, in part due to a much larger contributor basis in countries in the Global North as a consequence of socio-economic inequalities and the digital divide [5,6].

Albeit the manifold usage of OSM building footprints, an adequate investigation into their completeness on the global scale has not been conducted so far. This talk investigates OSM building completeness in regions home to a population of 3.5 billion people (about 50% of the global population). First, we propose a machine learning regression method based on Random Forests (RF) to assess OSM building completeness within all 13,189 urban centers (as defined by the European Commission [7]). The analysis utilizes an extensive collection of open building data from commercial and authoritative sources as training data and builds upon very recent technological advances to utilize OSM full-history data for spatio-temporal data analysis on the global scale [8]. The model further relies on information obtained from remote sensing data (land cover, population distribution, night time lights), subnational human development, and urban road network density as predictors. This allows us – for the first time – to present a comprehensive assessment of the evolution of urban OSM building completeness which encompasses all data contributed to OSM since 2008.

Herfort, B., Lautenbach, S., Porto de Albuquerque, J., Anderson, J., & Zipf, A. (2022). Inequalities in the completeness of OpenStreetMap buildings in urban centers.
In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.). Proceedings of the Academic Track at State of the Map 2022, Florence, Italy, 19-21 August 2022. Available at https://zenodo.org/communities/sotm-22
DOI: 10.5281/zenodo.7004534

 

For each urban center we calculated the OSM building completeness using the area ratio method which has been applied by several other researchers in the context of urban areas [9–11]. Several measures have been adopted to describe the temporal evolution of inequality in urban OSM building mapping on the global scale and per World Bank region. We report on the average monthly OSM building completeness for urban centers globally and distinguished this score further by World Bank region and SHDI. These analyses has been conducted for annual snapshots from 2008-01-01 up until 2022-01-01.

We investigated the impact of humanitarian mapping through the HOT Tasking Manager and corporate mapping by Apple, Meta, MapBox, Microsoft and Kaart on overall completeness and inequality measures. OSM contributions have been considered as humanitarian mapping activities following the approach developed by Herfort et al. which utilizes information obtained from a HOT Tasking Manager database dump [12]. Corporate mapping activities were identified by OSM user ID, expanding on the approach presented in [13] by using a mapper's self-disclosed corporate affiliation in their OSM user bio instead of relying on out-of-date lists on the OSM wiki [14]. Based on this information, we derived the share of humanitarian map edits and corporate map edits on the overall OSM building data.

Overall, average urban OSM building completeness is estimated at 21% globally. Relatively high completeness was estimated for Europe & Central Asia (67%) as well as for North America (56%). Completeness values lower than the global average were observed for the regions Latin America & Caribbean (17%), East Asia & Pacific (16%), Middle East & North Africa (11%), and South Asia (7%). The completeness value for East Asia & Pacific was strongly influenced by the fact that urban centers in China were hardly mapped, very likely because mapping in OSM is prohibited by law. Sub-Saharan Africa completeness (29%) was slightly higher than the global mean. These regional differences illustrate that the global average is of limited explanatory power.

Distinguishing urban centers by SHDI also revealed dramatic differences in the temporal trajectories of completeness. In general, urban centers in regions with very high SHDI had the highest levels of mapped building completeness. Surprisingly, however, there was no positive correlation between SHDI and completeness. The completeness in low SHDI urban centers was higher than the completeness of urban centers with high SHDI. Our results suggest that this was due to the positive impact of organized humanitarian mapping activities since 2015, especially on urban centers located in low and medium SHDI regions.

We found that organized humanitarian mapping activities in urban centers contributed an average of about 8% of the building footprints globally. However, humanitarian contributions were focused on specific regions, especially in Africa where about 43% of all building edits in Sub-Saharan Africa were related to organized humanitarian mapping activities. Overall, organized humanitarian mapping activities were expectedly associated with lower subnational human development index values, in line with previous findings [12]. We generally found corporate mapping activity to constitute less than 2% of all building edits globally (and only about 0.1% in urban centers), a significant difference in participation from corporate mappers editing nearly 20% of the global road network as previously found [13].

The results reveal that for 1,510 cities home to a population of more than 400 Million people, OSM building footprint data is more than 80% complete and can provide an alternative to otherwise complex approaches utilized to derive authoritative and/or automated building datasets. The digital divide in OSM has receded over the past decade, but still exists. As such, OSM data completeness improved, but was still strongly biased by regional, socio-economic and demographic factors on several scales. This echoes the highly uneven geographies of participation observed in Wikipedia [15] and stands in contrast to the relatively higher and more evenly distributed completeness for OSM's road network [16]. If this trend continues, OSM will become more complete, but will still not evolve towards a truly global inclusive map. As a consequence, global studies and global frameworks (such as SDGs) which use OSM data will draw wrong conclusions and will provide misleading recommendations for decision makers when the biases in OSM's coverage are not accounted for.

The results reveal the need to address the remaining stark data inequalities, which could not be turned around so far by humanitarian and corporate organized mapping activities. We conclude with recommendations directed at stakeholders working with OSM data: (1) Multi-scale building completeness measures should be applied before subsequent usage of OSM data to outline the potential negative effect of missing data. (2) Completeness maps should be used in combination with socio-demographic information to guide future mapping activities to ensure that "nobody is left behind" as encouraged by the SDGs.

All Python code and Jupyter notebooks necessary to calculate the geospatial statistics, create maps and derive figures are available in this GitHub repository: https://github.com/GIScience/global-urban-building-completeness-analysis.

## References

[1] Oort, P. (2006). *Spatial data quality: from description to application*. PhD thesis, Wageningen Universiteit.

[2] Neis, P., Zielstra, D. & Zipf, A. (2011). The Street Network Evolution of Crowdsourced Maps: OpenStreetMap in Germany 2007–2011. *Future Internet, 4,* 1–21.

[3] Barron, C., Neis, P. & Zipf, A. (2014). A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis. *Transactions GIS, 18*, 877–895.

[4] Senaratne, H., Mobasheri, A., Ali, A.L., Capineri, C. & Haklay, M. (2016). A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science, 8816*, 1–29.

[5] Sui, D., Goodchild, M. & Elwood, S. (2013). Volunteered geographic information, the exaflood, and the growing digital divide. In: Sui, D., Elwood, S., & Goodchild, M. (Eds.) *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice,* Springer, Dordrecht, 1–12.

[6] Neis, P., Zielstra, D. & Zipf, A. (2013). Comparison of Volunteered Geographic Information Data Contributions and Community Development for Selected World Regions. *Future Internet, 5,* 282–300.

[7] Florczyk, A., Corbane, C,. Schiavina, M., Pesaresi, M., Maffenini, L., Melchiorri, M., Politis, P., Sabo, F., Freire, S., Ehrlich, D., Kemper, T., Tommasi, P., Airaghi, D., & Zanchetta, L. (2019). GHS Urban Centre Database 2015, multitemporal and multidimensional attributes. *European Commission, Joint Research Centre*. Retrieved from https://ghsl.jrc.ec.europa.eu/ghs_stat_ucdb2015mt_r2019a.php

[8] Raifer, M., Troilo, R., Kowatsch, F., Auer, M., Loos, L., Marx, S., Przybill, K., Fendrich, S., Mocnik, F.B., & Zipf, A. (2019). OSHDB: a framework for spatio-temporal analysis of OpenStreetMap history data. *Open Geospatial Data, Software and Standards*, *4*(1), 1–12.

[9] Fan, H., Zipf, A., Fu, Q., & Neis, P. (2014). Quality assessment for building footprints data on OpenStreetMap. *International Journal of Geographical Information Science, 28*, 700–719.

[10] Törnros, T., Dorn, H., Hahmann, S. & Zipf, A. (2015). Uncertainties of completeness measures in OpenStreetMap–A case study for buildings in a medium-sized German city. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, II-3/W5*, 353–357.

[11] Hecht, R., Kunze, C. & Hahmann, S. (2013). Measuring Completeness of Building Footprints in OpenStreetMap over Space and Time. *ISPRS International Journal of Geo-Information, 2,* 1066–1091.
[12] Herfort, B., Lautenbach, S., Porto de Albuquerque, J., Anderson, J. & Zipf, A. (2021). The evolution of humanitarian mapping within the OpenStreetMap community. *Scientific Reports, 11*(1), 1-15.
[13] Anderson, J., Sarkar, D. & Palen, L. (2019). Corporate Editors in the Evolving Landscape of OpenStreetMap. *ISPRS International Journal of Geo-Information, 8*, 232.
[14] OpenStreetMap Wiki (2022). Organised Editing/Activities. Retrieved from https://wiki.openstreetmap.org/wiki/Organised_Editing/Activities
[15] Graham, M., Straumann, R.K. & Hogan, B. (2015). Digital Divisions of Labor and Informational Magnetism: Mapping Participation in Wikipedia. *Annals of the American Association of Geographers, 105,* 1158–1178.
[16] Barrington-Leigh, C., & Millard-Ball, A. (2017). The world's user-generated road map is more than 80% complete. *PLOS ONE, 12*(8), e0180698.

# The cell size issue in OpenStreetMap data quality parameter analyses: an interpolation-based approach

Silvana Camboim[1,*], Caio Paiva[1], Marcio Schmidt[1], Elias Elias[1], Everton Bortolini[1], Guilherme Neivas[1]

[1] Programa de Pós Graduação em Ciências Geodésicas,Federal University of Parana, Curitiba, Brazil; silvanacamboim@gmail.com, anjospaiva@ufpr.br, marcio.schmidt@ufu.br, elias_naim2008@hotmail.com, evertonbortolini@ufpr.br, guilhermesneivas@gmail.com

* Author to whom correspondence should be addressed.

This abstract was accepted to the Academic Track of the State of the Map 2022 Conference after peer-review.

Knowing the quality of a given geospatial data allows measuring how much its use can be viable in specific applications and assist in decision making. ISO 19157 [1] established that the geospatial data quality indicators are positional accuracy, temporal accuracy, thematic accuracy, logical consistency, and completeness. These measures are represented by values that summarize the condition of a product as a whole. These values tend to be homogeneous throughout the evaluated area in traditional mapping. In contrast, in VGI, data quality can be affected by several conditions related to editing history, contribution period, and contributor profiles [2,3]. Given the mentioned aspects, data quality in VGI platforms tends to be heterogeneous, i.e., the results may show significant discrepancies according to the area assessed or even within the same region.

Given the heterogeneity issues described, several researchers around the world have performed the quality assessment of these types of information based on the principle of subdividing the study area into cells [4–8]. Such a procedure has been used in extrinsic quality assessment processes based on ISO 19157 indicators or intrinsic parameters associated with the characteristics of the contributions and contributors. Given the results obtained, the representation of the quality of the data from sub-areas makes it possible to obtain analyses regarding the existence of patterns and establish relationships with other agents and their predominance. The discretization of space into rectangular or hexagonal grids is central to this type of analysis.

The subdivision can occur using grids or based on different sizes of areas. The units with non-equal area cells allow us to perform analyses accepting other features or spatial phenomena that define these dimensions (e.g., neighbourhood border, a river or a railway track, areas with different population densities, and the dichotomy between rural and urban areas). However, these methods make operations difficult because they demand that the area value weigh the values; and the spatial analysis considering the neighbourhood is more complex. Units with equal area cells (e.g. grid with rectangular or hexagonal cells) solve

Camboim, S., Paiva, C., Schmidt, M., Elias, E., Bortolini, E., & Neivas, G. (2022). The cell size issue in OpenStreetMap data quality parameter analyses: an interpolation-based approach.
In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.). Proceedings of the Academic Track at State of the Map 2022, Florence, Italy, 19-21 August 2022. Available at https://zenodo.org/communities/sotm-22
DOI: 10.5281/zenodo.7004550

 

these two limitations. However, the problem of the grid of cells not conforming to spatial phenomena or features reappears. In order to conform to them, it is necessary to determine the optimal size of the cells.

However, one issue remains little discussed: how to determine the size of such cells. Using too large a cell would treat unequal areas equally. On the other hand, using too small a cell and the increased computational cost of the process, ultimately, the ability to generalize the results is lost. Therefore, in this work, we seek to develop an interactive approach for determining the grid cell size calculation, initially using points of known positional accuracy. The hypothesis here is that when the analyzed subarea is of optimal size, one can interpolate the error within the cell via an IDW and generate minimal residuals at the control points. Furthermore, by consecutively subdividing the grid, the mean squared error versus cell size curve will approach stability, thus revealing the optimal size for a given region.

IDW interpolation calculates cell values using sample point sets. This method considers that the higher weights in the interpolation should be due to the proximity of the unknown value point. Thus, if we had a homogeneous behaviour of the quality parameter in an individual area, by interpolation, we could estimate the quality of the points where this value was unknown.

The methodological procedures developed using python in the QGIS environment are:

For the study area, points of known positional accuracy are chosen (in our case, intersections of the road system), from which a random subset of 10% is separated as a control set.

- Definition of a first grid;
- The points are used for interpolation within each cell by the IDW method. The Root-mean-square deviation (RMSE) is calculated using the control points for each cell and the average of the RMSEs for the entire area;
- Definition of a second grid with half the resolution of the first grid;
- Repeat the process described in item 3 for the second grid;
- Calculate the differences between the average error values of the second grid and the first grid and check their significance;
- Repetition of the process described in case there are still values indicated as significant.

In a first analysis, we did a preliminary study for a Brazilian city, Curitiba, with about 28 thousand points of known accuracy. We separated 2.8 thousand control points, and the city was divided into 8 km to 250 m cells. From the preliminary study performed, it was noted that the method show promise in obtaining the necessary analyses to identify the aspects proposed in this work. Furthermore, it was noticed that, as the cell size decreased, the results tended to be more constant, which corroborates the hypothesis of this relationship with data quality. The next steps are to continue the analyses, starting with the verifications and the representation of the magnitude of the differences between different cell sizes.

Although it is a method that still has a relatively high computational cost to be realized, the results are exciting and can be optimized. It is assumed that if it is possible to identify the minimum cell size in which it is possible to estimate the quality of the features, this will help in decision making regarding the incorporation of procedures in different areas. This method may need even smaller clippings in regions with very heterogeneous characteristics concerning their surroundings (e.g., slums). It is an initial approach to resolve with data a fundamental issue arising from the lack of knowledge of the granularity of discrepancies for each study area.

## References

[1] International Organization for Standardization (2013). *ISO 19157:2013 Geographic Information – Data quality*. Retrieved from https://www.iso.org/standard/32575.html

[2] Barron, C., Neis, P., & Zipf, A. (2014). A comprehensive framework for intrinsic OpenStreetMap quality analysis. *Transactions in GIS, 18*(6), 877–895.

[3] Minghini, M., & Frassinelli, F. (2019). OpenStreetMap history for intrinsic quality assessment: Is OSM up-to-date?. *Open Geospatial Data, Software and Standards, 4*(1), 1–17.

[4] Haklay, M., Basiouka, S., Antoniou, V., & Ather, A. (2010). How many volunteers does it take to map an area well? The validity of Linus' law to volunteered geographic information. *The Cartographic Journal, 47*(4), 315–322.

[5] Fan, H., Zipf, A., Fu, Q., & Neis, P. (2014). Quality assessment for building footprints data on OpenStreetMap. *International Journal of Geographical Information Science, 28*(4), 700–719.

[6] Martini, A., Kuper, P.V., & Breunig, M. (2019). Database-supported change analysis and quality evaluation of OpenStreetMap Data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 4*, 535–541.

[7] Brovelli, M.A., & Zamboni, G. (2018). A new method for the assessment of spatial accuracy and completeness of OpenStreetMap building footprints. *ISPRS International Journal of Geo-Information, 7*(8), 289.

[8] Zhang, H., & Malczewski, J. (2017). Accuracy evaluation of the Canadian OpenStreetMap road networks. *International Journal of Geospatial and Environmental Research, 5*(2).

# Investigating the capability of UAV imagery for AI-assisted mapping of Refugee Camps in East Africa

Christopher Yan-Chak Chan[1,*], Matthias Weigand[2], Emran Alchikh Alnajar[3], Hannes Taubenböck[1,2]

[1] Department of Remote Sensing, Julius Maximilians University of Würzburg, Würzburg, Germany; chrischank@protonmail.ch, yan_chak_christopher.chan@stud-mail.uni-wuerzburg.de
[2] German Remote Sensing Data Center (DFD), German Aerospace Center (DLR), Weßling, Germany; matthias.weigand@dlr.de, hannes.taubenboeck@dlr.de
[3] Humanitarian OpenStreetMap Team, Washington D.C., USA; omran.najjar@hotosm.org

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2022 Conference after peer-review.

Refugee camps and informal settlements provide accommodation to some of the most vulnerable populations, with many of them located in Sub-Saharan Africa. Many of these settlements lack up-to-date geoinformation that we take for granted in developed settlements. Having up-to-date maps on the dimension, spatial layout are important for assisting administration tasks such as crisis intervention, infrastructure development, and population estimates which encourage economic productivity [1]. The data inequality between the developed and developing countries are results of multiple reasons ranging from a lack of commercial interest to knowledge gaps in data contributors [2–4]. Such disparity can be reduced using assisted mapping technology. A combination of Very High Resolution (VHR) satellite imagery and Machine Learning (ML) based methods that exploit the textural, spectral, and morphological characteristics of VHR imagery are commonly used to extract geospatial and imagery of these complex environments [5]. Although many have shown promising results in satellite VHR scenarios (e.g. differentiating slum and non-slum [6,7]), in drone imagery, however, results might suffer from noise caused by environment and drone-based specific motion problems. Recent advances in Deep Learning (DL) based Computer Vision (CV), however, might be able to address these issues [8].

In our pilot project, we investigated the capabilities of applying DL semantic segmentation methods for delineating building footprints in refugee camps from open-data drone imagery. This study is connected to a larger initiative to open-source the AI assisted mapping platform in the current Humanitarian OpenStreetMap Team's (HOT) ecosystem. The study focuses on two refugee camps in East Africa located in a similar savannah ecosystem. The first camp is located in Dzaleka, Dowa, Malawi, this area is split into the Dzaleka North and Dzaleka main camp, home to ca. 40,000 refugees. The northern camp is characterised by newer, spatially well-planned metal-sheeted roofs, while the southern main

Chan, C.Y., Weigand, M., Alnajar, E.A., & Taubenböck, H. (2022). Investigating the capability of UAV imagery in AI-assisted mapping of Refugee Camps in East Africa.
In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.). Proceedings of the Academic Track at State of the Map 2022, Florence, Italy, 19-21 August 2022. Available at https://zenodo.org/communities/sotm-22
DOI: 10.5281/zenodo.7004576

camp is characterised by arrangements of dense mud-walled buildings with stone-lined thatched-roofs [9]. The second camp, the Kalobeyei settlement is home to ca. 34,800 refugees as of 2019, the settlement part of an extension for the larger Kakuma refugee camp, located in the rural county of Turkana, North-West Kenya. This camp is significantly more spacious and is characterised by spatially well-planned shelters with metal-sheeted roofs [10]. VHR drone images were provided for both camps by the OpenAerialMap project and vector labels produced by HOT volunteers were provided for the Dzaleka camps, and vector labels for the Kalobeyei settlements were created specifically for this study.

DL and semantic segmentation applications aim at classifying each pixel of an image into a predefined set of semantic classes. Semantic segmentation image classification tasks require high quality of reference data. In particular, the motion artefact problem [8] needs to be addressed. However, in the case of this study, a large quantity of available labels from the Dzaleka camps were not created with these challenges in mind. These imperfections in labelling could cause the trained model to misclassify. In order to have a model which performs well on drone imagery, we hypothesise that this will be a significant feature for the model to learn. Therefore, the Kalobeyei dataset was labelled to be the pixel-aligned dataset accounting for motion artefacts. Through this unique combination of different data quality levels, their influence on the classification results could be tested.

Several architectures of Convolutional Neural Networks (herein CNN) have recently developed. From these, the U-Net architecture [11] was selected on the basis of applicable ability in many domains, a proven track record of performance in remote sensing segmentation tasks, and relative computational efficiency [12–14]. The symmetrical encoder-decoder type architecture is able to extract deeper features in the encoder layers, while the decoder layers recover and interpolate spatial features [15]. The ability to switch out the encoder structure allows the DL practitioner to experiment with more up-to-date architectures without changing the output shape. This drastically increases the combination of experiments that allow testing the best combinations of encoder-decoder structure suitable for the dataset. All the experiments in this study were carried out using the high-level PyTorch API Segmentation-Model-PyTorch [16]. In this study, the experiments with changed encoder came from the EfficientNet family. There are three reasons for this selection: Firstly, at one of the last stages of the Open-Cities-AI-Challenge (OCC) competition winning network [17], EfficientNet B1 was used as an encoder. Secondly, the EfficientNet family is a set of network architectures that are structured and easy to scale up when computational resources become available. Thirdly, they are an accepted representation of generalised state-of-the-art architectures that have been tested and performed well in classical CV datasets [18]. In essence, these are sets of experiments that mix and match old and new architectural design. Regarding data pre-processing, the drone imagery was normalised and resampled to 15 cm resolution, then subsequently cropped at two-thirds overlapping steps to increase the quantity of training data. The training, validation, and testing dataset were split on a 60, 30, and 10% ratio. With appropriate augmentation techniques applied to the training and validation dataset.

The objective of this study is to test the U-Net and several variations of the architectures' performance for building footprint mapping, initially on the pixel-aligned and simple Kalobeyei dataset. Subsequently, the models were introduced to the less-aligned Dzaleka and Dzaleka North datasets of higher complexity. A comparison of the baseline experiments, which kept the settings of hyperparameters to be consistent, were conducted between the architectures. Class-based accuracy assessments metrics were used to evaluate performances between the baseline experiments setup. This allows evaluation into which level of data quality is required to achieve acceptable classification results when scaling DL assisted mapping efforts in the context of humanitarian mapping.

Initial baseline experiments suggest limited transferability from the competition winning OCC model. This indicates that the OCC model might be over-generalised to the competition test dataset, which mainly consists of a similar drone imagery of 10 different urban areas in Africa. This is accentuated by this model achieving very high confidence on metal-sheeted roofs, while not detecting any of the more complicated thatched roofs common in the Dzaleka camp (see Figure 1, left). The OCC model was found to also struggle with some of the more obscure drone motion artefacts occurring at the edge of the imagery in the Kalobeyei camp.

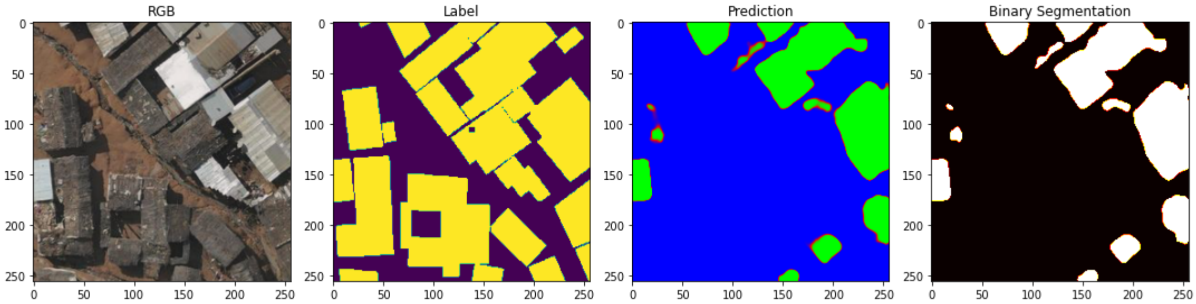


*Figure 1*. Prediction example from further training of EfficientNet B1 U-Net of the OCC winning network.

In most of our experiments, precision and recall measures have both reached above 0.7. However, there are still significant variations among different architectures and between the datasets. Precision and recall suggest that unmodified U-Nets were least affected by the introduction of the less accurate samples from the Dzaleka dataset. Meanwhile, in absolute terms, the EfficientNet encoder U-Nets performed better if only the Kalobeyei dataset is considered. The results show that deeper versions of the respective network architecture had not universally produced improvements, but they vary by dataset. Segmentation results from the OCC model (see Figure 1), when adapted to the settlements of this study, did not outperform the other tested architectures.

This study demonstrated the ability to use DL semantic segmentation to perform building segmentation in complex humanitarian applications. Having increased access to open-data VHR drone imagery such as the OpenAerialMap initiative is an advantage to building AI-assisted humanitarian mapping. The study evaluated various U-Net based architectures and data input setup. Yet, the variation of the results not only emphasised the complexity of DL based methods, but also indicate that further efforts will be needed to focus on the selection of suitable network architectures as they show varying resilience towards imperfect reference samples.

## References

[1] Herfort, B., Lautenbach, S., de Albuquerque, J.P., Anderson, J., & Zipf, A. (2021). The evolution of humanitarian mapping within the OpenStreetMap community. *Scientific Reports, 11*(1), 1–15.

[2] Anderson, J., Sarkar, D., & Palen, L. (2019). Corporate Editors in the Evolving Landscape of OpenStreetMap. *ISPRS International Journal of Geo-Information, 8*(5), 232.

[3] Veselovsky, V., Sarkar, D., Anderson, J., & Soden, R. (2021). An automated approach to identifying corporate editing activity in OpenStreetMap. In: Minghini, M., Ludwig, C., Anderson, J., Mooney, P., & Grinberger, A.Y. (Eds.) *Proceedings of the Academic Track at the State of the Map 2021 Online Conference*, 31–33.

[4] Yang, A., Fan, H., Jing, N., Sun, Y., & Zipf, A. (2016). Temporal Analysis on Contribution Inequality in OpenStreetMap: A Comparative Study for Four Countries. *ISPRS International Journal of Geo-Information, 5*(1), 5.

[5] Taubenböck, H., Kraff, N.J., & Wurm, M. (2018). The morphology of the Arrival City—A global categorization based on literature surveys and remotely sensed data. *Applied Geography, 92*, 150–167.

[6] Kuffer, M., Pfeffer, K., Sliuzas, R., & Baud, I. (2016). Extraction of Slum Areas From VHR Imagery Using GLCM Variance. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9*(5), 1830–1840.

[7] Wurm, M., Weigand, M., Schmitt, A., Geiß, C., & Taubenböck, H. (2017). Exploitation of textural and morphological image features in Sentinel-2A data for slum mapping. *2017 Joint Urban Remote Sensing Event (JURSE)*, 1–4.

[8] Smith, M.W., Carrivick, J.L., & Quincey, D.J. (2016). Structure from motion photogrammetry in physical geography. *Progress in Physical Geography: Earth and Environment, 40*(2), 247–275.

[9] UNHCR (2014). Malawi: Joint Assessment Mission Report. Dzaleka Refugee Camp (November 2014). Retrieved from https://www.unhcr.org/protection/health/5680f7d09/malawi-joint-assessment-mission-report-dzaleka-refugee-camp-november-2014.html

[10] UNHCR & Ministry of Foreign Affairs of Denmark (2019). UNHCR/Danida - The Integrated Solutions Model in and Around Kalobeyei, Turkana, Kenya. Retrieved from https://www.unhcr.org/research/evalreports/5dfa287c4/unhcrdanida-integrated-solutions-model-around-kalobeyei-turkana-kenya.html

[11] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W.M., & Frangi, A.F. (Eds.) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015,* Springer, Cham, 234–241.

[12] Hoeser, T., Bachofer, F., & Kuenzer, C. (2020). Object Detection and Image Segmentation with Deep Learning on Earth Observation Data: A Review—Part II: Applications. *Remote Sensing, 12*(18), 3053.

[13] Hoeser, T., & Kuenzer, C. (2020). Object Detection and Image Segmentation with Deep Learning on Earth Observation Data: A Review-Part I: Evolution and Recent Trends. *Remote Sensing, 12*(10), 1667.

[14] Xu, Y., Wu, L., Xie, Z., & Chen, Z. (2018). Building Extraction in Very High Resolution Remote Sensing Imagery Using Deep Learning and Guided Filters. *Remote Sensing, 10*(1), 144.

[15] Wurm, M., Stark, T., Zhu, X.X., Weigand, M., & Taubenböck, H. (2019). Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing, 150*, 59–69.

[16] Yakubovyskiy, P. (2021). Segmentation Models documentation. Retrieved from https://segmentation-models-pytorch.readthedocs.io/en/latest/index.html

[17] Open Cities AI Challenge: Segmenting Buildings for Disaster Resilience. (2022). Retrieved from https://github.com/drivendataorg/open-cities-ai-challenge

[18] Tan, M., & Le, Q.V. (2020). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv preprint*, arXiv:1905.11946.

# Corporate editing and its impact on network navigability within OpenStreetMap

Corey Dickinson[1,*], Jugal Patel[1] and Dipto Sarkar[2]

[1] Department of Geography, McGill University, Canada; corey.dickinson@mail.mcgill.ca, jugal.patel@mail.mcgill.ca
[2] Department of Geography and Environmental Studies; Carelton University, Canada; dipto.sarkar@carleton.ca

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2022 Conference after peer-review.

OpenStreetMap (OSM) contributors have traditionally lacked explicit monetary incentives for contribution [1]. Since 2016, several corporations (including Apple, Facebook, Microsoft, and Uber) have increasingly contributed data to OSM. Corporate editors (CEs) represent a distinct community as their editors are compensated and thus their contributions cannot be labeled as 'volunteered'. Additionally, corporations employ large editing teams and new state-of-the-art editing techniques aided by artificial intelligence, making them capable of editing large swaths of information in relatively short time [2]. Corporate teams are often led by long-time OSM community members themselves, emphasizing the multifaceted nature of a rapidly growing open mapping platform [3]. While there has been some contention about the quality of edits done by CEs, corporations argue their contributions improve existing data [4]. Our study provides a preliminary quantitative evaluation of data quality impacts of corporate edits on OSM.

We assess intrinsic data quality across five regions that have high levels of corporate contributions: Dallas-Ft. Worth, Egypt, Jamaica, Thailand, and Singapore. The quality of these regions is compared to that of Denmark, a region which has witnessed relatively less corporate interest, yet possesses a well-mapped OSM presence due to a well developed local mapping community [5]. These evaluations were performed using measures of intrinsic map quality. While the most straightforward evaluation methods involve comparing against extrinsic sources, such as either ground reference information or authoritative data sources; lack of data availability, licensing terms, and costs often render this comparison untenable [4,6,7]. A transferrable, data driven way of assessing quality remains using Intrinsic Quality Indicators (IQIs), a sub-field of OSM analysis which provides a variety of approaches for evaluating intrinsic OSM data quality. We chose to focus on IQIs that apply to networks, and to evaluate IQIs for land-based transportation networks within OSM. We analyzed networks for our specified locations for every other year between 2014 and 2022.

OSM editing archives were processed using R to extract maps of the relative activity of corporate editors [8]. This was done in order to evaluate the intensity of corporate edits by region over time, and to select regions with high rates of organized editing. Our list of

Dickinson, C., Patel, J., & Sarkar, D. (2022). Corporate editing and its impact on network navigability within OpenStreetMap.
In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.). Proceedings of the Academic Track at State of the Map 2022, Florence, Italy, 19-21 August 2022. Available at https://zenodo.org/communities/sotm-22
DOI: 10.5281/zenodo.7004583

 

corporate editors was sourced from OSM's publicly available list of corporate editors accounts. We extracted entire networks that represented the first day of each year of interest (2014, 2016, 2018, 2020, 2022) from OSM's historical archives. For the purposes of this study, we extracted all networks composed of all OSM ways having *highway* as a key, which represents the entire road network.

We evaluated several IQIs for our areas of interest. We focused on completeness of network, both in terms growth over time and in terms of its navigability. We operationalized "completeness for navigability" as an intrinsic measurement by exploring the percentages of networks that possessed attributes necessary for GPS navigation – street names and speed limits. Navigability was assessed and compared across time points using Origin-Destination matrices. By creating a regular matrix across the area and calculating the ratio between a direct route between points, and a route navigated within our network, we calculated a ratio that can be compared across time to evaluate the changing efficiency of the navigable network. Additionally, when building routing networks, we discovered an additional IQI : the presence and qualities of topological islands within our network. That is, areas which are disconnected from the main network due to mapping errors or incompleteness.

After mapping these metrics, we analyzed how they correlate with each other and how they change over time. This analysis consisted of both a general comparison of patterns across years and regions, and more detailed analysis of statistical trends in completeness of networks. Overall, IQI trends for the road network reveal consistent patterns across all measures and locations. There is a trend towards increasing data quality in terms of gradual increase of network length, completeness in terms of attributes (name, speed limit, and pedestrian access), the increasing efficiency of Origin-Destination Matrix routing ratios, and the increasing amount of places that have "navigable" attributes. Importantly, we found differences between our control location (Denmark) and our other areas of interest. The primary difference of note is not with regards to the quality of the data, but with respect to the rate at which data quality improves: Denmark's rate of quality improvement is slower than other locations. The faster rate of quality improvement in the test areas highlights that the data creation and editing activity by corporate editors and other organized editors in these locations are helping narrow gaps in data quality.

While this presentation highlights the trends of data quality increase, it does not tease apart the quality assessment of contributions by corporate teams versus other mapping groups. As a crowdsourcing platform, data in OSM is co-produced by repeated editing of data objects by different members of the community [9]. The appearance of CEs in OSM represents the arrival of another community of 'producers' in the OSM ecosystem, and thus a new evolution in its overall trajectory [2,10]. Consequently, there is significant interaction between CEs and non-CEs in data co-production in OSM, further reinforcing the idea that OSM is a 'community of communities' [11]. Each location has their own patterns regarding editing communities, what they edit, and the sociopolitical and economic ground truth in the real world. Each of these factors impact the data, and may make comparing editing patterns difficult, especially given the diversity of motivations both with CE communities and within other OSM communities. Hence, we do not try to pry apart the differences in trends between individual countries. Instead, we focus on the overall trend between our test and control locations. With these caveats, we find that the quality of the network has increased in these areas across all tracked metrics at a faster rate than it has in areas with low rates of corporate edits, indicating that corporate editing may have a positive effect on the overall quality of the map.

## References

[1] Budhathoki, N.R., & Haythornthwaite, C. (2013). Motivation for open collaboration: Crowd and community models and the case of OpenStreetMap. *American Behavioral Scientist*, *57*(5), 548–575.
[2] Anderson, J., Sarkar, D., & Palen, L. (2019). Corporate editors in the evolving landscape of OpenStreetMap. *ISPRS International Journal of Geo-Information, 8*(5), 232.
[3] Dickinson, C. (2021) Inside the 'Wikipedia of Maps,' tensions grow over corporate influence. Retrieved from https://www.bloomberg.com/news/articles/2021-02-19/openstreetmap-charts-a-controversial-new-direction
[4] Antoniou, V., & Skopeliti, A. (2015). Measures and indicators of VGI quality: an overview. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, II-3/W5*, 345–351.
[5] Coast, S. (2015). *The Book of OSM*. CreateSpace Independent Publishing Platform, California, USA.
[6] Barron, C., Neis, P., & Zipf, A. (2014). A comprehensive framework for intrinsic OpenStreetMap quality analysis. *Transactions in GIS*, *18*(6), 877–895.
[7] Senaratne, H., Mobasheri, A., Ali, A. L., Capineri, C., & Haklay, M. (2017). A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science*, *31*(1), 139–167.
[8] R Foundation (2022) The R Project for Statistical Computing. Retrieved from https://www.R-project.org
[9] Mooney, P., & Corcoran, P. (2014). Analysis of Interaction and Co-editing Patterns amongst OpenStreetMap Contributors. *Transactions in GIS*, *18*(5), 633–659.
[10] Solís, P. (2017). Building mappers not just maps: challenges and opportunities from YouthMappers on scaling up the crowd in crowd-sourced open mapping for development. *AAG annual meeting. American Association of Geographers,* Boston, Massachusetts, 7 April 2017.
[11] Sarkar, D., & Anderson, J.T. (2022). Corporate editors in OpenStreetMap: Investigating co-editing patterns. *Transactions in GIS*, *6*, 1879–1897.

# Returning the favor - Leveraging quality insights of OpenStreetMap-based land-use/land-cover multi-label modeling to the community

Moritz Schott[1,*], Adina Zell[2], Sven Lautenbach[1,3], Begüm Demir[2] and Alexander Zipf[1,3]

[1] Institute of Geography/GIScience, Heidelberg University, Heidelberg, Germany; moritz.schott@uni-heidelberg.de, sven.lautenbach@uni-heidelberg.de, zipf@uni-heidelberg.de
[2] Faculty of Electrical Engineering and Computer Science/Remote Sensing Image Analysis Group, Technische Universität Berlin, Berlin, Germany; adina.zell@campus.tu-berlin.de, demir@tu-berlin.de
[3] HeiGIT, Heidelberg University, Heidelberg, Germany

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2022 Conference after peer-review.

Land-use and land-cover (LULC) information in OSM is a challenging topic. On the one hand, this information provides the background for all other data rendered on the central map and is used by applications like https://osmlanduse.org. It has a high potential to be used as a valuable data source to tackle major current challenges like the climate crisis. On the other hand, this information has a difficult position within the OSM ecosystem. LULC information can be quite cumbersome or even difficult to map e.g. due to natural ambiguity. As most other OSM tagging schemes, the current LULC tagging scheme is the result of a bottom-up growth which resulted in a collection of sometimes ambiguous, unstable or overlapping tag definitions that are not fully compatible with any official LULC legend definition [1]. Furthermore, the data is highly shaped by local or national preferences and imports. This diversity of the LULC data in OSM is a fundamental principle of OSM that enabled the success of the project. Yet, this can create considerable usage barriers or at least caveats for data users unfamiliar with the projects' ecosystem. The remote sensing community for instance has started to use OSM LULC information as labels in their classification models. Frequently, OSM LULC data has thereby been taken at face value without critical reflection. And, while the quality and fitness for purpose of OSM data has been proven in many cases (e.g., [2,3]), these analyses have also unveiled quality variations e.g. between rural and urban regions. The quality of OSM therefore can be assumed to be generally high, but remains unknown for a specific use-case.

The proposed work first assesses the impact of these challenges on a use-case of multi-label remote sensing (RS) image classification and then provides a machine learning (ML) based workflow to overcome and finally mitigate them. RS images contain multiple

Schott, M., Zell, A., Lautenbach, S., Demir, B., & Zipf, A. (2022). Returning the favor - Leveraging quality insights of OpenStreetMap-based land-use/land-cover multi-label modeling to the community.
In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.). Proceedings of the Academic Track at State of the Map 2022, Florence, Italy, 19-21 August 2022. Available at https://zenodo.org/communities/sotm-22
DOI: 10.5281/zenodo.7004593
 

LULC classes and thus can be simultaneously associated with different class labels (i.e. multi-labels). The development of multi-label RS image classification methods, which aim at automatically assigning class labels to each RS image in an archive, is a growing research interest. In the presented study multi-labels are extracted from OSM and used to train the ML model.

The fitness for purpose of OSM for multi-label RS image classification was tested on a Sentinel 2 scene with a resolution of 10m and four bands in south west Germany recorded in June 2021. The area was chosen for its estimated high completeness of OSM land use information and for the low amount of imported data. OSM data was grouped by its tags into the four LULC classes 'forest', 'agricultural area', 'build-up area' and 'water body'. 18 tags that could unequivocally be mapped to these classes were used and small elements, below the image resolution or the classes minimal mapping unit, were filtered. The chosen scene was then tiled into a 1.22 x 1.22 km grid of 8100 image patches. Zero to four labels were assigned to each patch based on the OSM LULC elements therein. Evaluation was performed manually on 910 random patches, of which 80% had a correct OSM-based multi-label, thereby proving the assumed high completeness and quality in the region.

The proposed workflow provides a method to enhance this OSM-based RS image multi-label classification and extend it to areas of lower OSM quality and completeness using ML, specifically deep learning (DL). The main obstacle for ML and especially DL is the required amount of labeled training data. Volunteered geographic information (VGI) like OSM offers a potential solution to this challenge by providing an overabundance of LULC information that is suitable for this purpose - if data quality is sufficiently high. The workflow uses the multi-label information extracted from OSM for training and then detects discrepancies between its predictions and OSM. Using this information and pinpointing the exact location of error within the patches provides valuable OSM data quality information.

Apart from facilitating a fast quality estimation for large areas, the workflow can also make its findings automatically available to the OSM community in a feedback loop using the HOT Tasking Manager framework. Thereby, the valuable service by the OSM community of providing large amounts of free and generally high quality training data is recognised in the form of quality feedback including mapping hints to the OSM community. The five workflow stages are: 1) RS data collection and preprocessing, 2) OSM data collection and preprocessing, 3) LULC multi-label modeling, 4) OSM data issue flagging and 5) the community feedback loop via Tasking Manager projects. While each step is an atomic use case and application, the combination of all four steps creates a tool that is useful for the RS and the OSM community likewise. The tool is openly available at [https://gitlab.gistools.geog.uni-heidelberg.de/giscience/ideal-vgi/osm-multitag](https://gitlab.gistools.geog.uni-heidelberg.de/giscience/ideal-vgi/osm-multitag) under the GNU Affero General Public License v3 including example datasets. Manual input was kept as low as possible while enabling the 'human in the loop' to take full control over all input and output.

The workflow extracts multi-label training data in stages 1) and 2) as described. Stage 3) then trains a DL model to: i) predict multi-labels; ii) identify incorrect (training) labels; and iii) localise the areas associated with these incorrect labels. Multi-labels can be incorrect if either one or more labels are missing (omission) or labels are wrongly assigned (commission). Omission errors occur when data is missing in OSM while commission errors occur when OSM data is falsely mapped in the tile. Multi-labels that are only partly incorrect, e.g. containing only a single wrong label, are often called ‘noisy labels’.

A DeepLabv3+ model [4] where the segmentation head is replaced with a multi-label classification head [5] is used as classification model. Yet, the output of the model does not provide spatial information regarding the location of the LULC classes present in the images. To localise the classes in a given image, we investigate the effectiveness of explainable neural networks. Explanation methods are capable of generating explanations in the form of pixel-level heatmaps that can be highly relevant for providing class localisation maps from a DL model trained using image level labels. In this work, we exploit self-enhancement maps [6] due to its proven success in providing accurate explanations. The class localisation is then combined with error detection techniques allowing the visualisation of potentially incorrect labels from the OSM data in stage 4). Depending on the availability of verified noise-free labels, one of two different approaches for noise detection is applied. If no clean data is available, the model class prediction is directly compared with the corresponding label from OSM. If noise-free data is available, an adapted noise detection method of CleanNet [7] for multi-labels is applied. CleanNet generates a single representative class prototype for each class and uses it to estimate the correctness of sample labels. Thereby, the model is first trained on clean data. Afterwards, for each class, a prototype is extracted based on the feature maps of the noise-free data. To detect if an image has wrong labels, the extracted features of the image with potentially noisy labels are compared with the corresponding class prototypes. The similarity between the two indicates whether the OSM label is potentially noisy.

For demonstration, the model was trained on the described Sentinel 2 scene in Germany. The models' performance was validated on the manually labelled 910 patches. The model achieved a mean average precision (MAP) of 0.98, whereas the direct use of OSM tags yields a MAP of 0.91. For the selected test region, the OSM data quality was very high and the above described filter procedure assured a balanced amount of omission and commission errors. To assess the effectiveness of the model under higher label-noise rates, simulating areas of lower OSM quality or completeness, synthetic noise was added to the labels using the approach established by [8]. This ensures that both omission and commission label noise are introduced to the multi-label training set. When synthetically decreasing the OSM training data quality to a MAP of 0.78, the model maintained a MAP above 0.97. Models trained on further deteriorated training data were also able to achieve surprisingly high prediction rates. Detailed investigation though showed that these models had moved towards a probabilistic approach of predicting classes more and more based on their occurrence in the training set rather than the actual image features.

The experiment shows that for DL approaches, label quantity is often of higher importance compared to label quality. Given the sheer overabundance of OSM training labels, OSM data quality can therefore be seen as a secondary problem, especially regarding the fact that it is generally high and can partly be assessed and assured through prior data processing. It can also be stated that OSM noise detection using DL is possible even for models trained on areas with relatively low OSM quality and completeness. Yet, the separation of stages 3) and 4) also allows the application of pretrained models in case OSM data is completely missing or suspected to be below a usable quality threshold. While this application is under active research, it is suspected that pretrained models from regions with higher OSM data quality can yield good multi-label prediction and noise detection results in regions with very low OSM quality and completeness. Provided the existence of a comparable region in terms of class distribution, definition and appearance, this would make the workflow widely applicable to most global regions or historic timestamps, independent of local OSM data.

The final stage 5) uses the above described localised potential OSM data errors to create HOT Tasking Manager projects via the API. These projects provide additional correction hints. Yet, no automatic editing takes place. The community is kept in full control of all mapping actions as a 'human in the loop'.

The high fitness for purpose of OSM has been proven for the use-case of multi-label RS image classification. The proposed tool provides an automated OSM multi-label extraction, modeling and verification procedure including a return of results to the OSM community. A major challenge of the approach is the tiled view on the data. If OSM assigns correct multi-labels to a patch, more fine grained data issues will not be detected. Yet, this approach allows large scale data assessments, before the topic of more detailed data improvement is tackled. It also allows to run repeated OSM LULC quality and completeness estimations for large areas over time including far reaching retrospectives. Another major benefit is the usage of local OSM data for modeling, thus making regionalised models the standard procedure. This is required for OSM LULC information as regional data structures and communities exist, which need to be preserved. The model can lead to regional homogenisation and data cohesion within these regional communities.

**Acknowledgements**: This work was funded by DFG, German Research Foundation ("IdealVGI", 424966858).

## References

[1] Fonte, C., Minghini, M., Antoniou, V., See, L., Patriarca, J., Brovelli, M.A. and Milcinski, G.(2016). An automated methodology for converting OSM data into a land use/cover map. In: Bandrova, T,. & Konecny, M. (Eds) *Proceedings of the 6th International Conference on Cartography & GIS*, Albena, Bulgaria, 13-17 June 2016, 482–473.

[2] Jokar Arsanjani, J., Mooney, P., Zipf, A., & Schauss, A. (2015). Quality Assessment of the contributed land use information from openstreetmap versus authoritative datasets. In: Jokar Arsanjani, J., Zipf, A., Mooney, P., & Helbich, M. (Eds.) *OpenStreetMap in GIScience*, Springer, Cham, 37–58.

[3] Fonte, C.C., Bastin, L., See, L., Foody, G., & Lupia, F. (2015). Usability of VGI for validation of land cover maps. *International Journal of Geographical Information Science, 29*(7), 1269–1291.

[4] Chen, L., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C., & Weiss, Y. (Eds.) *Computer Vision – ECCV 2018*, Springer, Cham, 833–851.

[5] Ahlswede, S., Madam, N.T., Schulz, C., Kleinschmit, B., & Demr, B. (2022). Weakly supervised semantic segmentation of remote sensing images for tree species classification based on explanation methods. *arXiv preprint*, arXiv:2201.07495.

[6] Zhang, X., Wei, Y., Yang, Y., & Wu, F. (2020). Rethinking Localization Map: Towards Accurate Object Perception with Self-Enhancement Maps. *arXiv preprint*, arXiv:2006.05220v2.

[7] Lee, K., He, X., Zhang, L., & Yang, L. (2018). CleanNet: Transfer Learning for Scalable Image Classifier Training with Label Noise. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5447–5456.

[8] Burgert, T., Ravanbakhsh, M., & Demir, B. (2022). On the effects of different types of label noise in multi-label remote sensing image classification. *arXiv preprint*, arXiv:2207.13975.

# Combining Volunteered Geographic Information and WPdx standards to Improve Mapping of Rural Water Infrastructure in Uganda

Stellamaris Nakacwa[1,*], Brent McCusker[1], Katherine Sill[2], Courtney Clark[3] and Rory Nealon[4]

[1] Department of Geology and Geography, West Virginia University, Morgantown, USA; sn00013@mix.wvu.edu, brmccusker@mix.wvu.edu
[2] Water Point Data Exchange, Global Water Challenge, Arlington, VA, USA; katy.sill@getf.org
[3] Everywhere She Maps, YouthMappers, Lubbock, TX, USA; cclark@americangeo.org
[4] USAID GeoCentre, DC, USA; rnealon@usaid.gov

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2022 Conference after peer-review.

Access to clean and safe drinking water is critical to public health and socioeconomic prosperity, yet an estimated quarter of the world's population lacks such [1]. This was evidenced by the unprecedented outbreak of the COVID-19 pandemic, which left communities extremely vulnerable to fatal illnesses due to the limited access to water for handwashing or lack of knowledge of the existence of the utility [2].

Subsequently, the lack of data on the distribution of drinking water resources poses a great challenge to optimizing water resource investments and has limited Artificial Intelligence/Machine Learning-enabled advancements in the water sector compared to all other sectors like heath [3]. Increasing the frequency of water point data collection and sharing through crowdsourcing and volunteered geographic information would greatly improve the availability of water point data, and contribute to the extended roles of water resource distribution, monitoring, and management, especially in rural communities. Therefore, this paper describes the methodology for combining different water mapping schemas to create comprehensive multi-platform drinking water infrastructure data and enhance rapid updates to support a suite of drinking water resource analytics, and extended advanced technology explorations towards improved decision-making.

Through the water infrastructure infrastructure mapping project in Gulu, Uganda, a comprehensive OpenStreetMap (OSM) water tag review was conducted to explicitly maximize the usage and application of these crowdsourced data in the Water Point Data Exchange (WPdx) [4] suite of analytical platforms for sharing, access and using data to improve decision making in the Water, Sanitation and Hygiene sector. A relative data model resulting from a region-specific knowledge landscape was developed to facilitate the

Nakacwa, S., McCusker, B., Sill, K., Clark, C., & Nealon, R. (2022). Combining Volunteered Geographic Information and WPdx standards to Improve Mapping of Rural Water Infrastructure in Uganda.
In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.). Proceedings of the Academic Track at State of the Map 2022, Florence, Italy, 19-21 August 2022. Available at https://zenodo.org/communities/sotm-22
DOI: 10.5281/zenodo.7004619
 

mapping of water infrastructure data points, and simultaneous updates of both the WPdx & OSM databases, through mobile data collection. Underlying the development of this data model/schema, a design criterion was established which guided and justified the overall selection of the most relevant factors to include in the process that would eventually become detailed to communicate water infrastructure and functionality for administrative decision support. The criteria were followed by an assessment of the compliance, consistency, completeness, and granularity [5] of the OSM tag to support the development of the OSM language in the Kobo toolbox [6]. A field mapping workflow was designed to facilitate the field-data collection employing the developed water infrastructure data model and Kobo toolbox.

More than 15,000 buildings, 1,400 $km^2$ of roads and over 500 water data points were added to OSM as well as the WPdx database for the later data. In addition, observations were made regarding the improvement of such processes and potential extension of such methods beyond one geographical area.

**References**

[1] WHO/UNICEF Joint Monitoring Programme for water supply, sanitation and hygiene (2021) Progress on household drinking water, sanitation and hygiene. Retrieved from https://washdata.org/sites/default/files/2022-01/jmp-2021-wash-households-highlights.pdf

[2] Howard, G., Bartram, J., Brocklehurst, C., Colford, J.M., Jr, Costa, F., Cunliffe, D., Dreibelbis, R., Eisenberg, J.N.S., Evans, B., Girones, R., Hrudey, S., Willetts, J., & Wright, C.Y. (2020). COVID-19: Urgent actions, critical reflections and future relevance of 'WaSH': lessons for the current and future pandemics. *Journal of Water and Health*, *18*(5), 613–630.

[3] Doorn, N. (2021). Artificial intelligence in the water domain: Opportunities for responsible use. *Science of The Total Environment*, *755*, 142561.

[4] Dickinson, N., Knipschild, F., Magara, P., & Kwizera, G. (2017). *Harnessing water point data to improve drinking water services*. Retrieved from https://nl.ircwash.org/sites/default/files/harnessing_waterpoint_data_to_improve_drinking_waterservices_-_white_paper.pdf

[5] Mocnik, F.-B., Zipf, A., & Raifer, M. (2017). The OpenStreetMap Folksonomy and Its Evolution. *Geo-Spatial Information Science*, *20*, 219–230.

[6] Lakshminarasimhappa, M.C. (2022). Web-Based and Smart Mobile App for Data Collection: Kobo Toolbox / Kobo Collect. *Journal of Indian Library Association*, *57*(2), 72–79.

# Floor plan extraction from digital building models

Helga Tauscher[1,]*, Subhashini Krishnakumar[1] and Dominik Heigener[1]

[1] Computing in Civil Engineering, Bauhaus-Universität Weimar, Weimar, Germany;
helga.tauscher@uni.weimar.de, subhashini.krishnakumar@uni-weimar.de,
dominik.heigener@uni-weimar.de

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2022 Conference after peer-review.

Official geodata is increasingly published not only in the form of 2D maps but also in 3D, mainly as city models in CityGML. Usually, the outer shell of buildings is captured in such models, but they may also involve more intricate details. Even more detailed building models are generated during the planning process for new buildings and renovations. Nowadays, owners and operators produce these models in digital form during the design, planning and execution. Then, an as-built version is archived for the remaining lifetime of a building. In the future, digital building model submission may even be required to obtain a building permit. At the same time, public interest in detailed information about public and semi-public interior spaces is increasing, for example about their accessibility, the localization of barriers, destinations (e.g., contact persons in public administration), resources (e.g., defibrillators) or to get a first impression in advance (e.g., virtual open day).

Since the context of creating and capturing geodata and building data is fundamentally different, there is hardly any integration. Indoor data for maps and navigation models is manually captured or at best derived in undocumented semi-automatic workflows. The project LevelOut sets out to develop automated methods and services to make detailed indoor data from digital building models selectively available for the population of city models, map and navigation services (in the form of 2.5D floor plans). Towards this end, we develop a platform to check building models for the required information, extract selected and simplified indoor data and convert it into various formats (see Figure 1). As conversion targets, we consider two application schemes of the Geographic Markup Language (GML), CityGML LoD0 Indoor [1] and IndoorGML [2], as well as the indoor portion of OpenStreetMap (OSM Indoor). As input, we rely on data in the format IFC (Industry Foundation Classes), the most widespread standardised data exchange format for digital building models. With this research and development, we aim to provide a workflow and software for systematic access of floor plan data in IFC [3,4]. We start at both ends of integration by looking at the detailed structures of the source and target models in parallel.

OSM Indoor is but one of the various extraction targets. We focus on geometry with Simple Indoor Tagging and additionally consider Simple 3D Buildings (S3DB) for basic tagging of 3D properties of the building. The data created could be directly fed into OSM or

Tauscher, H., Krishnakumar, S., & Heigener D. (2022). Floor plan extraction from digital building models.
In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.). Proceedings of the Academic Track at State of the Map 2022, Florence, Italy, 19-21 August 2022. Available at https://zenodo.org/communities/sotm-22
DOI: 10.5281/zenodo.7004634

 

may serve as a viable base for further mapping. There is only little work on IFC-to-OSM conversion. A recent attempt was made with the addition of IFC-functionality to the JOSM plugin 'Indoor Helper' [5]. The plugin is under active development, but it lacks a general approach on the IFC side as well as systematic coverage of the heterogeneous options to represent geometry in the IFC schema.

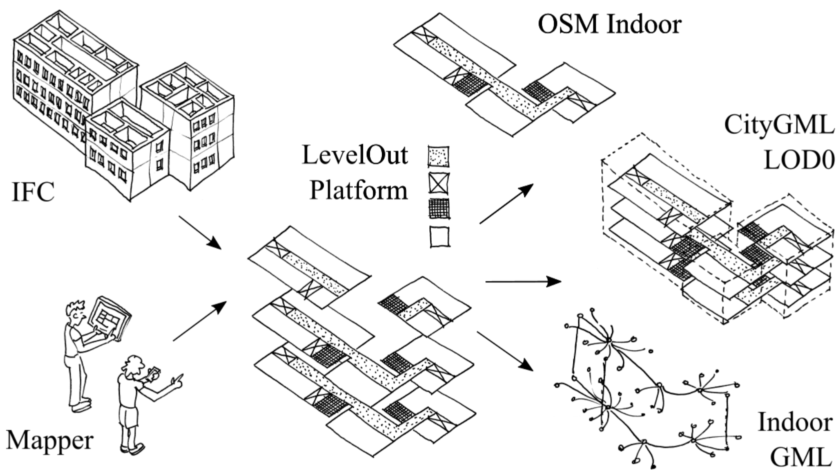

*Figure 1*. LevelOut Platform with input and output.

The second extraction target, CityGML, is a GML application schema for 3D city models and an exchange format issued as an international standard by the OGC. It provides a conceptual model for 3D city objects with their semantics, geometry, topology, appearance and other attributes. There are many different application scenarios for CityGML, including urban planning, simulations, real estate management, 3D cadastre, among others. There is a wide range of research on IFC-to-CityGML conversion. Konde et al. were first to put a focus on floor plans [6]. They demonstrate various options to model floor plans in CityGML based on inclusion of particular elements (walls, openings, rooms) and on choice of spatio-semantic representation (0-, 1-, 2-dimensional abstractions) resulting in different LoD. They also show two approaches to extract floor plans from IFC: either relying on the original IFC geometry or on preprocessed triangulated geometry.

IndoorGML is the third extraction target. The main purpose of this standard is to support navigation and location based services. It implements a cellular space model concept to represent geometry and topology. For the purpose of producing IndoorGML data from IFC, Diakite et al. recently published an open source solution [7]. Their tool carries out a conversion in three steps: parsing IFC data while organizing it as Linear Cell Complex (LCC) data structure, processing the data to generate IndoorGML cell boundaries, nodes, edges, and topological relationships and, finally, exporting the data as IndoorGML files.

In the current stage, we have already investigated and examined the OGC schemas (CityGML, IndoorGML) and analysed the OSM schema and tagging approaches. From the set of target models, we derive a common model that will have near-trivial mappings to OSM Indoor, CityGML and IndoorGML and serve as an intermediate model. In a subsequent step we are going to identify commonalities and shared characteristics across the three models in order to determine features to be included in the intermediate model. Preliminary considerations include geo-coordinates, rooms and space-defining surfaces and elements.

Next, we identify relevant information in the source model that can be used to establish the 2.5D geometry and topology of navigable spaces for the target models. IFC exposes a variety of geometry modelling constructions from CAD software, mainly following the modelling paradigm of constructive solid geometry (CSG). So far, we have identified the following principal representation options in IFC-conformant building models:

a) Direct floor plan representation in 2.5D: Here, we have 2D representations in 3D space, usually located at finished floor level for a particular storey. We distinguish two versions: space boundaries (floor and wall element footprints) versus abstract representations of space-defining elements (e.g. wall axis).

b) Extraction from CSG: Spaces and constructive elements are often represented as solids, which result from extrusion of a planar shape. If the solid has been extruded in the z-direction, its base shape and extrusion height can be used as a 2.5D representation.

c) Projection onto floor level: If the geometry is not in CSG form with extrusions, but in the form of BREP (boundary representation) or a mesh, then projection followed by a simplification of the projection result is a possible way to extract the 2.5D representation.

In contrast to OSM, where semantics are attached to geometry via tags, the IFC data model is founded on a reciprocal principle with geometry attached to a semantically typed object as an attribute. Associations between objects such as physical contact or spatial containment can possibly carry geometric attributes as well. Thus, while in OSM the geometric model elements are charged with meaning through tags, in IFC objects with meaning are equipped with geometric properties. We aim to transfer semantic meaning and attributes to the target models. Whether this is possible and how much effort it requires, will depend on the type of geometry and according extraction method mentioned above.

We have investigated the IFC standard for relevant entity types (classes) and give a brief overview in the following: For 2.5D floor plans, we need to process elements per building and storey and thus exploit the spatial hierarchy, an essential concept in IFC based on aggregation relationships as expressed with the class *IfcRelAggregates*. Considering the work's background in extracting floor plans for map services and navigation models, the geometry attributed to objects of the class *IfcSpace* represents the walkable space. For the geometric representation of *IfcSpace*, different concepts are provided in IFC corresponding to the representation principles above: 2D curves as footprints (a), swept - extruded and revolved - area solids (b), clipping with half space solids and boundary representation (c). Using extruded area solids as an example, the geometry is of class *IfcExtrudedAreaSolid*. An extrusion profile is then represented as an attribute of class *IfcArbitraryClosedProfileDef* or *IfcArbitraryProfileDefWithVoids* - if voids exist due to building elements in space such as columns. These profiles can be used as 2.5D floor plans. The geometry description is further placed in a local coordinate system with origin of class *IfcCartesianPoint* and axis directions represented with class *IfcDirection*. Besides *IfcSpace*, further classes are relevant for the work's purpose, e.g. elements for vertical navigation, i.a. *IfcStair*, *IfcDoor*, *IfcRamp* or *IfcTransportElement* as well as space-bounding elements, i.a. *IfcRelSpaceBoundary*.

At first, we check the given building models and extract the required entities and their attributes manually in order to derive automatic procedures later on. Furthermore, the relevant building elements of the demonstration object need to be exported losslessly from the authoring application into the IFC format.

After identification of the relevant entities, we are developing a three-stage process for the actual population of target models from IFC.

1. Building model enrichment: Information that can be represented in IFC will be reintegrated into the building model instead of being promoted to the intermediate model only.
2. Building to intermediate model: This is an essential step which will be covered with a flexible rule-based mapping.
3. Intermediate to target models: Following a careful design of the intermediate model, this step should be simple.

We will test the processes with data from public buildings, two sets of university campus buildings as well as one newly built municipal administration centre. From the assessment of the building data, we will also document modelling and export guidelines for BIM authoring software. As far as possible, the demo data will be made publicly available as open data. More importantly, the conversion procedures will be published as open source and a respective conversion service will be offered online.

In summary, our work provides practical benefit in terms of tools to support the mapping process as well as a scientific contribution in terms of spatial data integration and expert involvement via domain-specific languages. The practical benefit of the conversion is obvious: building owners can publish data of their publicly accessible spaces to aid volunteer mapping efforts. In the future, we also want to tackle updating, checking, and comparison with existing OSM indoor data. Scientific contributions are made in different ways: First, a generalised indoor model based on the OGC Indoor Feature Model is provided. The audience of this conference may find particularly interesting how OSM data fits with the generalised model. The generalised intermediate model will not only facilitate the integration of IFC with multiple targets besides OSM, but also the integration of OSM with multiple sources besides IFC. Further, we explore methods for flexible data integration with domain specialists and expert community involvement. To OSM this could be relevant in the light of the flexible and fluctuating tagging schemes requiring adjustments to conversion processes. Finally, but beyond the scope of this conference, we evaluate the applicability of bidirectional integration methods with multiple sources and targets via intermediary formats.

## References

[1] Open Geospatial Consortium (2021). *CityGML 3.0.0. OGC City Geography Markup Language (CityGML) Part 1: Conceptual Model Standard*. Retrieved from http://www.opengis.net/doc/IS/CityGML-1/3.0

[2] Li, K.-J. (2016). IndoorGML: A standard for indoor spatial modeling. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLI-B4*, 701–704.

[3] International Organization for Standardization (2018) *ISO 16739-1:2018 Industry Foundation Cases (IFC) for data sharing in the construction and facility management industries—Part 1: Data schema.* Retrieved from https://www.iso.org/standard/70303.html

[4] Borrmann, A., König, M., Koch, C., & Beetz, J. (Eds.) (2018). *Building Information Modeling: Technology Foundations and Industry Practice.* Springer, Cham.

[5] OpenStreetMap Wiki (2022) JOSM/Plugins/indoorhelper. Retrieved from https://wiki.openstreetmap.org/w/index.php?title=JOSM/Plugins/indoorhelper&oldid=2263317

[6] Konde, A., Tauscher, H., Biljecki, F., & Crawford, J. (2018). Floor plans in CityGML. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *IV-4/W6*, 25–32.

[7] Diakite, A., Díaz-Vilariño, L., Biljecki, F., Isikdag, Ü., Simmons, S., Li, K., & Zlatanova, S. (2022). ifc2indoorgml: An open-source tool for generating IndoorGML from IFC. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B4-2022*, 295–301.

# Understanding and modelling accessibility to public green in large urban centers using OpenStreetMap data

Alice Battiston[1,]* and Rossano Schifanella[1,2]

[1] Department of Computer Science, University of Turin, Turin, Italy; alice.battiston@unito.it, rossano.schifanella@unito.it
[2] ISI Foundation, Turin, Italy

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2022 Conference after peer-review.

As of 2020, around 55% of the worldwide population lives in urban areas and the World Bank estimates forecast an increase of around 1.5 times in the urban population by 2045. Cities are also major contributors to the climate-change, with a consumption of about 78% of the worldwide energy and a production of 60% of greenhouse gas emissions.

A transition toward greener cities is often called as one of the solutions to reduce the environmental impact of cities, but also to make the urban environments more liveable, with positive spillovers on the mental and physical health of their population. In this context, the United Nations' Sustainable Development Goals 11.7 [1] indicates the need to make cities more inclusive and safe, but also environmentally sustainable, calling for the universal provision of safe, inclusive, and accessible, green and public spaces. A proper evaluation of this target requires complementing standard average metrics, looking for instance at the surface of green areas per capita within an urban area, with more sophisticated metrics, that are able to capture the interplay between the spatial distribution of both the population and green areas within a city.

A few studies on selected cities worldwide highlighted the importance of considering this interplay [2–7]. Among these, a recent study on the city of Seoul [3] shows that vast portions of the parks in the city are located in outer areas so that frequent opportunities to visit them are relatively minimal. In general, urban green areas in Seoul are inadequately distributed in relation to population, land use, and development density. By contrast, in the case of Shanghai [6], the degree of accessibility to green areas appears to decrease as we move from the city core to the urban periphery. Using multiple regression and spatial lag regression analysis, the authors also found that housing price is negatively correlated with travel time to green space. This negative association translates directly into a large environmental inequality, wherein wealthier communities benefit more from green spaces than disadvantaged communities. A similar socio-economic, but also ethnic, stratification is observed in the city of Chicago, where white-majority census tracts generally enjoy a

Battiston, A., & Schifanella, R. (2022). Understanding and modelling accessibility to public green in large urban centers using OpenStreetMap data.
In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.). Proceedings of the Academic Track at State of the Map 2022, Florence, Italy, 19-21 August 2022. Available at https://zenodo.org/communities/sotm-22
DOI: 10.5281/zenodo.7004662

 

significantly higher degree of accessibility to green areas than minority-dominated census tracts [7]. The former ethnic group also presents a lower income-based green-areas accessibility inequity compared to the other racial-ethnic groups.

Efforts to move beyond case studies and provide more accurate cross-country indicators have led to the construction of the 'generalised potential access to green areas' from the European Commission, which is provided as one of the city-level indicators of the Global Human Settlement (GHS) - Urban Centers Database [8]. The metric measures the proportion of the urban population for urban centers included in the atlas living in high green areas. Based on satellite data on the Normalized Difference Vegetation Index, the metric is however agnostic with respect to the characteristics of these high green areas - for instance, whether these are public or private green areas - and any accessibility notion, since the metric does not consider that people can move from their residential location to nearby areas. These limitations are accounted for in a recent study for the European Environmental Agency [9], whose geographical coverage is however limited to specific urban hotspots in Europe, for which high-resolution land use data from the Urban Atlas [10] is available.

With its worldwide coverage and detailed mapping, the use of land use and street network data from OpenStreetMap (OSM) [11] allows to expand the analysis beyond the European boundaries. Our study provides a threefold contribution in this direction. First, we compare detailed high-resolution land use data on green uses for European hotspots included in the Urban Atlas [10] with green-related key-value pairs in OpenStreetMap for similar geographical areas. We use similarity indices to assess the degree of completeness of the OSM tags of natural land uses in urban environments and show how the quality varies according to the type of natural use and the size as well as the geographical area of the urban center under consideration. Second, we propose a framework for the monitoring of the target for large urban centers worldwide. In particular, by leveraging data from OSM and population estimates from the Global Human Settlement [12], we develop a framework to measure accessibility to public green in large urban centers worldwide at a high resolution. Our sample include more than 2500 cities, whose boundaries are extracted from the GHS Urban Center Database [8]. In particular, for each country, we include in our study the 50 largest urban centers by population, provided that the urban center has at least a population of 100.000 inhabitants. We identify natural green areas in each urban center using relevant OSM keys on *landuse*, *natural* and *leisure* (e.g.: *leisure=park*) and extract the walkable street network to measure walking distances using the OSM-based routing service Open Source Routing Machine (OSRM) [13]. Accessibility indices are then constructed for each populated cell of the population grid. Following the academic literature on urban accessibility, we build several accessibility indices, from a minimum distance index to exposure metrics. Figure 1A reports the walking distance (in minutes) to the closest public green area of at least three hectares from residential cells in the city center of Paris. The chart, instead, displays the proportion of the population with access to a green area of at least three hectares as a function of walking distance, for the cities of Tokyo, Paris, New York and Turin. The framework will also be also used to build an interactive tool to navigate our results, which can be customized to select the type of green of interest, as well as the size of the green area. The constructed indices and resulting database represent a valuable source of information for policymakers to identify cities that are missing out and direct attention to those subareas within otherwise well-performing cities where the degree of accessibility is still insufficient. The developed indices are then used to study the relationship between the measured level of accessibility and the structural characteristics of the cities and unveil the role of small green areas as accessibility enhancers, particularly in densely inhabited urban centers (Figure 1B, for the city of Paris). As a third contribution, we demonstrate how the framework can be used to simulate the impact of different urban interventions, from the addition of a new public green area to infrastructural interventions to the street network, to help policymakers to shape transitions toward more sustainable and accessible urban environments. Along this line, Figure 1C shows the impact of adding up to five new public green areas in the urban centers of Paris in terms of population (count and fraction) with access to a public green area of at least three hectares within 10 minutes from the residential location. The count reports the new population (in thousands) meeting the target, while the fraction is cumulative and include the entire population meeting the target for the different scenarios. The locations of the new public green areas are chosen to maximize the share of the population with access to green within the selected thresholds (10 minutes walking). The maps display in green residential cells meeting the target in the current scenario and after the implementation of the selected policy (addition of 1, 2 or 5 new public green areas). All Python code and materials developed for this project are made available at the GitHub repository: [https://github.com/alibatti/AccessToGreenOSM](https://github.com/alibatti/AccessToGreenOSM).

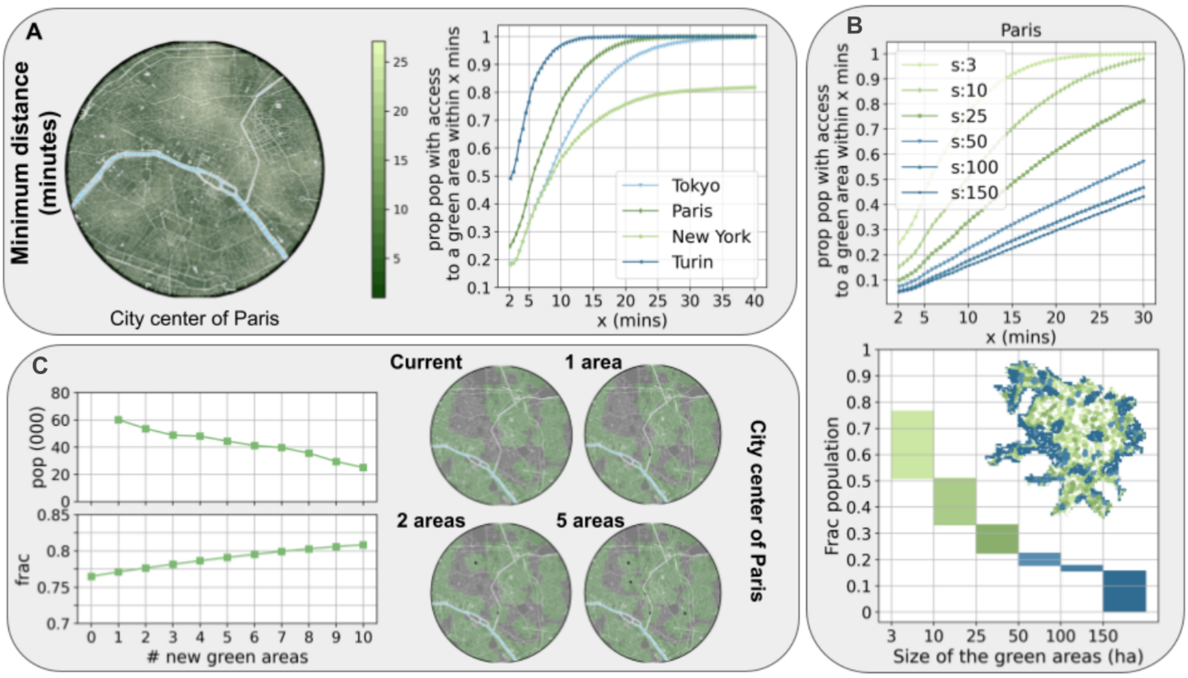


Figure 1. A: Walking distance (in minutes) to the closest public green area of at least three hectares for the city center of Paris (map) and the proportion of the population with access to a green area of at least three hectares as a function of walking distance, for the cities of Tokyo, Paris, New York and Turin (chart). B: The proportion of the population with access to a green area of at least s hectares as a function of walking distance in Paris, for increasing minimum park size thresholds s (top). The contribution of public green areas of different size to the overall fraction of the population with access to a public green area of at least 3 hectares within 10 minutes from their residential location (bottom). C: The impact of adding up to five new public green areas in Paris.

## References

[1] Lee, B.X., Kjaerulf, F., Turner, S., Cohen, L., Donnelly, P.D., Muggah, R., Davis, R., Realini, A., Kieselbach, B., MacGregor, L.S., Waller, I., Gordon, R., Moloney-Kitts, M., Lee, G., & Gilligan, J. (2016). Transforming our world: implementing the 2030 agenda through sustainable development goal indicators. *Journal of public health policy, 37*, 13–31.
[2] Giuliani, G., Petri, E., Interwies, E., Vysna, V., Guigoz, Y., Ray, N., & Dickie, I. (2021). Modelling accessibility to urban green areas using Open Earth Observations Data: A novel approach to support the urban SDG in four European cities. *Remote Sensing, 13*(3), 422.
[3] Oh, K., & Jeong, S. (2007). Assessing the spatial distribution of urban parks using gis. *Landscape and urban planning, 82*(1-2), 25–32.
[4] Rahman, K.A., & Zhang, D. (2018). Analyzing the level of accessibility of public urban green spaces to different socially vulnerable groups of people. *Sustainability, 10*(11), 3917.
[5] Zhang, J., Yue, W., Fan, P., & Gao, J. (2021). Measuring the accessibility of public green spaces in urban areas using web map services. *Applied Geography, 126*, 102381.
[6] Chen, Y., Yue, W., & La Rosa, D. (2020). Which communities have better accessibility to green space? an investigation into environmental inequality using big data. *Landscape and Urban Planning, 204*, 103919.
[7] Liu, D., Kwan, M.-P., & Kan, Z. (2021). Analysis of urban green space accessibility and distribution inequity in the city of chicago. *Urban Forestry & Urban Greening, 59*, 127029.
[8] Florczyk, A., Corbane, C,. Schiavina, M., Pesaresi, M., Maffenini, L., Melchiorri, M., Politis, P., Sabo, F., Freire, S., Ehrlich, D., Kemper, T., Tommasi, P., Airaghi, D., & Zanchetta, L. (2019). GHS Urban Centre Database 2015, multitemporal and multidimensional attributes. *European Commission, Joint Research Centre*. Retrieved from https://ghsl.jrc.ec.europa.eu/ghs_stat_ucdb2015mt_r2019a.php
[9] Poelmann, H. (2020). A walk to the park? Assessing access to green areas in Europe's cities. *European Commission Working Papers*, *WP 01/2018*.
[10] European Environment Agency (2020). Urban Atlas 2018. *Copernicus Land Monitoring Service*. Retrieved from https://land.copernicus.eu/local/urban-atlas
[11] OpenStreetMap contributors (2022). Planet OSM. Retrieved from https://planet.osm.org
[12] Schiavina, M., Freire, S., & MacManus, K. (2019). GHS population grid multi-temporal (1975, 1990, 2000, 2015) R2019A. *European Commission, Joint Research Centre*. Retrieved from https://data.jrc.ec.europa.eu/dataset/ed8e8e11-62c3-4895-a7b9-5ef851f112ed
[13] Luxen, D., & Vetter, C. (2011). Real-time routing with OpenStreetMap data. In: *Proceedings of the 19th ACM SIGPATIAL International Conference on Advances in Geographic Information Systems*, Chicago, IL, 1-4 November 2011, 513–516.

# Leveraging OpenStreetMap to investigate urban accessibility and safety of visually impaired pedestrians

Achituv Cohen[1], Asya Natapov[2] and Sagi Dalyot[3,*]

[1] Department of Geography, University of California, Santa Barbara, USA; achituvcohen@ucsb.edu
[2] School of Architecture, Building and Civil Engineering, Loughborough University, UK; a.natapov@lboro.ac.uk
[3] Mapping and Geoinformation Engineering, The Technion, Haifa, Israel; dalyot@technion.ac.il

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2022 Conference after peer-review.

Many efforts that include city policy and planning strategies are implemented to encourage urban active mobility. The outcome of these actions is measured by how transportable and accessible the city is. Although contribution is evident, in practice, the commonly used measures mostly disregard a huge part of the population that have mobility impairments requiring specific accessibility needs and preventing them to be an equal part of the sustainable city vision [1,2].

This research suggests using OpenStreetMap (OSM) data in customized analytical models to assess the accessibility level of the urban environment for visually impaired pedestrians. The models analyze the city on two levels: routing and accessibility. These are evaluated, correspondingly, based on possible routes, e.g., how long the optimal route is for visually impaired pedestrians compared to the shortest one, and on area, e.g., what is the overall accessibility and safety of a predefined urban extent. The play of both measures enables us to quantify the level of mobility and accessibility of the analyzed city.

To do so, we implement the following steps:

1. We examine the navigation preferences of visually impaired pedestrians in the urban space. This allows a better understanding of the various environmental and morphological factors and characteristics of the urban form that promote safe and accessible navigation [3,4]. These are translated into spatial and temporal criterion: a) Way Type, which quantifies how suitable the path is in terms of usage and safety; b) the existence of Vision Impairment Assistive Landmarks that support safe wayfinding and navigation; c) Way Complexity, which measures the level of linearity of the path; and d) Crowdedness, which measures the overall pedestrian traffic volume.
2. We transform OSM's street network into a weighted graph, where for each graph edge we calculate the cost according to the above criteria. Cost is derived from

Cohen, A., Natapov, A., & Dalyot, S. (2022). Leveraging OpenStreetMap to investigate urban accessibility and safety of visually impaired pedestrians.
In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.). Proceedings of the Academic Track at State of the Map 2022, Florence, Italy, 19-21 August 2022. Available at https://zenodo.org/communities/sotm-22
DOI: 10.5281/zenodo.7004666

 

segments (and OSM points and polygons nearby) that facilitate safe and accessible walking for visually impaired pedestrians (e.g., separated sidewalks and straight paths), and segments that hinder safe and accessible walking for visually impaired pedestrians (e.g., shared and overcrowded streets).

3. We develop three analytical models that measure the accessibility level of the urban environment for visually impaired pedestrians: a) street-based, which relies on averaging the costs of all graph edges for a given area, hence it can be implemented for different urban levels (spatial extents); b) centrality-based, which adds on the street-based the centrality indices betweenness and closeness [5] that consider the significance of each graph edge in the street network in respect to all other edges (high centrality values mostly signify streets that attract large pedestrian traffic flow); c) route-oriented method, in which numerous routes are generated on the graph for location tuples, defined by graph vertices, and then the ratio of the optimal route for visually impaired pedestrians and the shortest route (commonly used for seeing pedestrians) is evaluated. The smaller the ratio value, the more accessible the route.

The developed models are evaluated for Greater London, the UK. 33 boroughs with their wards are analyzed, resulting in processing 421,107 streets, 377,164 OSM nodes and 634,871 OSM ways. Results show the existence and spatial distribution of accessibility problems for visually impaired pedestrians. The street-based model, depicted in Figure 1, highlights the fact that in some urban areas, nature and green spaces, which are typically considered as contributing to wellbeing and encourage walking, are less accessible for visually impaired people, mostly due to the existing road types, e.g., gravel and dirt roads or shared spaces (bikes and pedestrians that share the same path), which are less accessible for this population. The centrality-based model shows that central streets are mostly more accessible, meaning that borough centers and connected streets are considered in general as accessible.

The route-based model, where more than 1,500,000 routes (with length shorter than 1,000 meters) were calculated, showed that on average the optimized routes are 11% longer and 17.5% more accessible than the shortest ones. Some optimal walking routes are twice as long as the shortest ones, where some impose safety issues that critically endanger visually impaired pedestrians. Wards that have a large proportion of street segments with poor accessibility evenly distributed throughout the ward tend to show less efficient route planning in terms of optimal routes that are considerably longer. In general, the route-based model produces clearer results to understanding the city's morphology in terms of accessibility for visually impaired pedestrians.

To a large extent, these models depend on the quality of OSM data, such as feature completeness and tag correctness. In terms of completeness, we found that sidewalks and crossings, which are two important model features, are not always mapped in OSM, mostly in the outskirts of London. One solution is to use learning methods and prediction models to complete missing data. In terms of tag correctness, we found that some inconsistencies exist with certain tags. One solution can be to make tag definitions in, e.g., OSM Wiki, more inclusive and clear, with a focus on accessibility aspects.

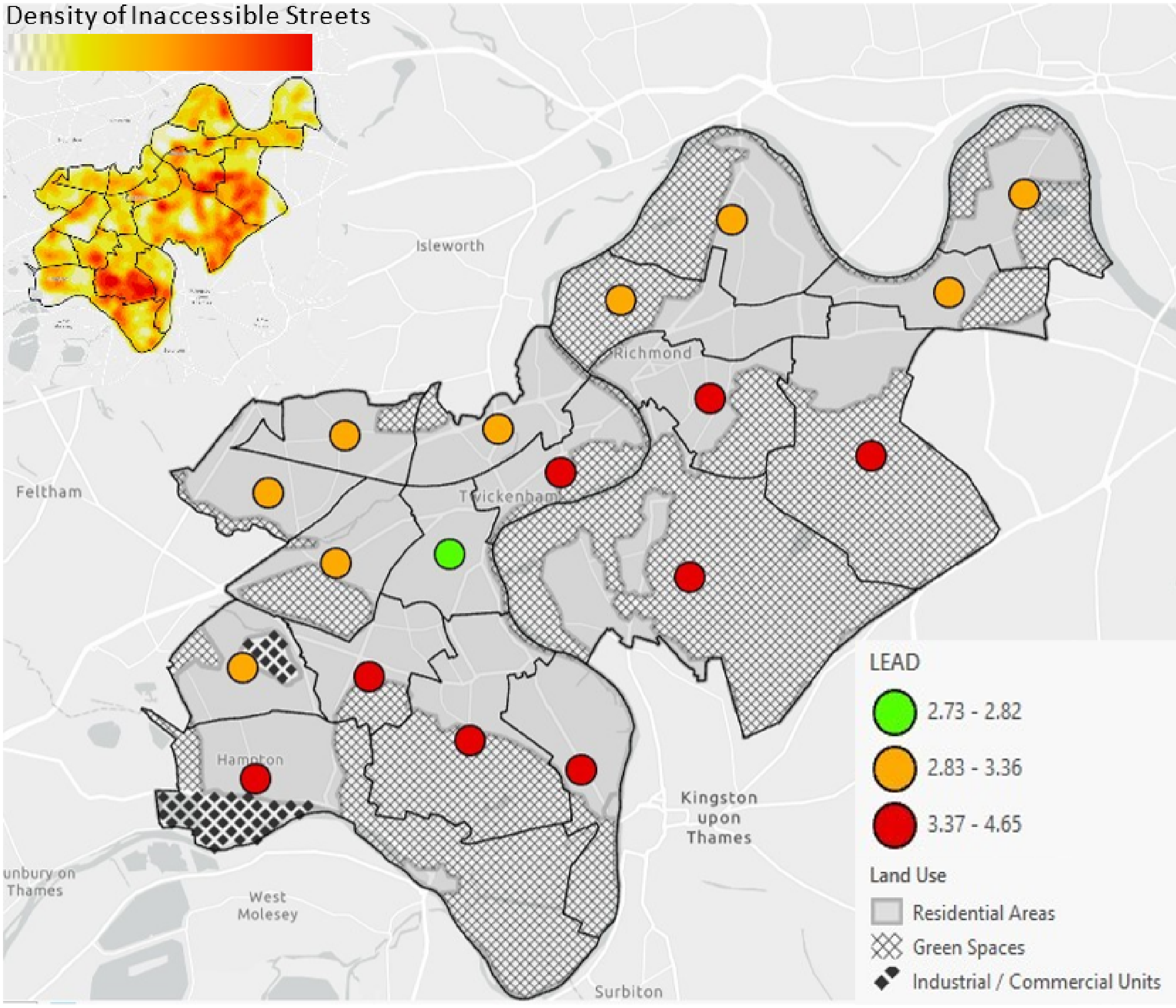


*Figure 1*. Richmond upon Thames spatial extent and its wards' boundaries (black polygons). The spatial distribution of land uses , and wards' LEAD values (denoted by colored circles). Top Left - Heat map of inaccessible street segments (in red).

Results of this study show how various accessibility levels for visually impaired pedestrians might be assessed and where they are found in the city, pointing to the existing problems this community faces when navigating. These include challenging street network connectivity and dangerous walking areas. Based on this work, the urban accessibility for other imparied communities can be considered while adjusting the navigation preferences and criterion used in the developed analytical models. Moreover, a more fine resolution of accessibility can also be considered, e.g., accessibility to shops and buildings.

The results demonstrate that the current practice of urban planning and design worldwide still suffers from lack of democratization, limiting the mobility and navigation of certain groups. The accessibility models developed in this research can be used for better city planning and design, enhancing the city mobility and walkability equality and improving quality of life for these vulnerable road users. Our findings provide analytical tools to enable decision makers, city stakeholders and practitioners to enrich management, monitoring and development of their cities, and support sustainable, livable lifestyles and walkability equality. These, in turn, will ease navigation and mobility of visually impaired pedestrians, overall improving health outcomes and their integration into society.

## References

[1] Campisi, T., Ignaccolo, M., Inturri, G., Tesoriere, G., & Torrisi, V. (2021). Evaluation of walkability and mobility requirements of visually impaired people in urban spaces. *Research in Transportation Business & Management, 40*, 100592.
[2] Kocaman, S., & Ozdemir, N. (2020). Improvement of disability rights via Geographic Information Science. *Sustainability, 12*(14), 5807.
[3] Cohen, A., & Dalyot, S. (2021). Route planning for blind pedestrians using OpenStreetMap. *Environment and Planning B: Urban Analytics and City Science, 48*(6), 1511–1526.
[4] Cohen, A., & Dalyot, S. (2020). Machine-learning prediction models for pedestrian traffic flow levels: Towards optimizing walking routes for blind pedestrians. *Transactions in GIS, 24*(5), 1264–1279.
[5] Cooper, C.H.V. (2015). Spatial localization of closeness and betweenness measures: a self-contradictory but useful form of network analysis. *International Journal of Geographical Information Science, 29*(8), 1293–1309.

# OpenStreetMap as a tool for skill building

Aishworya Shrestha[1,*], Nama Raj Budhathoki[1] and Nancy Erbstein[2]

[1] PEER Science Project, Kathmandu Living Labs, Kathmandu, Nepal; aishworya.shrestha@kathmandulivinglabs.org, nama.budhathoki@kathmandulivinglabs.org
[2] Associate Vice Provost, Global Education; Associate Professor In Residence, School of Education, University of California, Davis, USA; nerbstein@ucdavis.edu

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2022 Conference after peer-review.

OpenStreetMap, the crowdsourced geospatial database, currently has over eight million registered members [1]. This makes it one of the largest VGI projects with proven multifaceted use cases, e.g. post-disaster response, combating female genital mutilation, app development, and navigation. Its contributors wholly make and maintain the database, making all decisions without a top-down governing authority. Generally, OSM mapping is regarded as a form of volunteering to create freely accessible geodata. However, recent studies suggest that experience gained through the mapping process could be equally important as well [2–4]. Building on the existing body of knowledge, in this talk, we will share the findings of our research on the effects of OSM mapping on young mappers.

We studied a youth mapping internship called Digital Internship and Leadership (DIAL) conducted in three cohorts. The cohorts were composed of 3 males and 5 females in cohort I, 12 males and 7 females in cohort II and 8 males and 2 females in cohort III. They were originally from different places but were living in Kathmandu and Pokhara. They were university level students studying business administration, crisis management, architecture, public health, computer science, and geomatics engineering. We chose this internship program for its inclusiveness in terms of academia, gender, and the geographical locations that the participants came from. The program was designed and executed by Kathmandu Living Labs (KLL) and funded by the National Academy of Sciences under the PEER Science Project. It was designed to fill critical map data gaps in rural Nepal through youths. Participant mappers were called through an open invitation on social media. They filled an application through an online form, which comprised questions regarding backgrounds and interests. The short-listed recent high school graduates and undergraduate students were from diverse academic backgrounds (geomatics engineering, architecture, crisis management, management, forestry, computer science and engineering, electronics engineering, management, public health, mechanical engineering). The participant mappers were trained in person for 3-4 days. They were oriented on remote mapping (Java OpenStreetMap and iD Editor), field mapping (OSM Tracker for Android, Field Papers), and geospatial data visualization (Carto) during the orientation.

We studied the self-assessed experiences of the participant mappers at two different points of time: (i) during the mapping program (ii) after two years for Cohorts II and

Shrestha, A., Budhathoki, N.R., & Erbstein, N. (2022). OpenStreetMap as a tool for skill building.
In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.). Proceedings of the Academic Track at State of the Map 2022, Florence, Italy, 19-21 August 2022. Available at https://zenodo.org/communities/sotm-22
DOI: 10.5281/zenodo.7004677

 

III, and three years for Cohort I. Short-term effects were studied through grounded theory coding of reports and blogs documented by all the participants during the internship period. Each participant had prepared at least one blog and one report each to document their experiences of mapping. These reports were collected from the participant mappers at the end of the internship and the blogs were written throughout the internship. The reports were written in a fixed format, whereas the blogs were flexible. The format of the report comprised an introduction to digital internship, process, outcome, and evaluation of the impact of their work, reflections on the personal and professional learning and growth emerging from the internship, and recommendations. For long-term impacts, an online survey administered to identify if the effects persisted. An online questionnaire was administered via Google forms to assess the extent to which effects reported during the mapping period persisted and whether others emerged. Questions were derived from the literature and based on analyses of the reports and blogs. The survey assessed learning in areas of affective learning, production of geography-based knowledge, networking and social skills development, technical skills and digital literacy, professional skill-building, cognitive skills and civic skills. All the participant mappers of DIAL were approached for the survey. The questionnaire was sent through email and google form and was self-administered. The response rate was 72.5%. First, we used descriptive statistics to summarise and describe the answers to each question, this consisted of calculating the frequency and mode of each data. The outliers were calculated through the interquartile range. The distinction in academia between geography, geomatics, and the others and its plausible effect on the long-term impacts were tracked through multivariate analysis.

Results show OSM mapping helps the mappers develop several vital skill sets and expand their knowledge in a variety of areas. Some of them are: deepening of civic engagement, development of social identity, expansion of geographic knowledge, spatial awareness, increase in happiness and satisfaction. They retain most of these skills even in the long run, irrespective of differences in academic or professional backgrounds. Surprisingly, 44.8% of the participants cited they considered being a professional mapper or cartographer at some point in time because of their experience in DIAL. The same people report that OSM mapping increased their belief in their ability to help society.

Our findings build upon the studies of the use of OSM in high schools, which was noted to increase creativity and spatial awareness among the students [2,3,5]. When compared to Ebrahim et al.'s study with ten-year-olds [3], the similarity in findings suggests these developments might be similar across ages. These developments suggest new directions toward the use of OSM as a tool for youth skill building and youth community engagement and design newer incentive mechanism for people to join and retain in OSM. 55.2% of the survey participants said that they continued to map in OSM after the internship. Reasons for discontinuation of mapping included absence of peer group, lack of time, lack of incentive, lack of interest and absence of recognition. Reasons for continuation to map included professional alignment to mapping, humanitarianism, social responsibility, interest to map, useability of the OSM data among others.

There is still a huge scope of investigation left in this area, ideally through a longitudinal study with a bigger and more diverse sample and comparisons between different program designs, to fully understand the wide array of effects of OSM mapping on the mappers, as well as the potential to deepen positive outcomes via associated youth learning and leadership programs. There are undoubtedly other categories of benefits of

OSM mapping that is yet to be identified. Hence, it is worthwhile to reconsider the idea of participatory mapping and related programs, and their effects on the contributing mappers.

## References

[1] Neis, P. (2022). Statistics of the free wiki world map (OepnStreetMap.org). Retrieved from https://osmstats.neis-one.org

[2] Coetzee, S., Minghini, M., Solís, P., Rautenbach, V., & Green, C. (2018). Towards Understanding The Impact Of Mapathons; Reflecting On Youthmappers Experiences. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-4/W8*, 35–42.

[3] Ebrahim, M., Minghini, M., Molinari, M.E., & Torrebruno, A. (2016). Minimapathon - Mapping The World At 10 Years Old. In: *Proceedings of the 8th Annual International Conference on Education and New Learning Technologies (EDULEARN 2016)*, Barcelona, Spain, 4-6 July 2016, 4200–4208.

[4] Hite, R., Solís, P., Wargo, L., & Larsen, T. (2018). Exploring Affective Dimensions of Authentic Geographic Education Using a Qualitative Document Analysis of Students' YouthMappers Blogs. *Education Sciences*, *8*(4), 173.

[5] Price, M., Berdnyk, A., & Brown, S. (2019). Open Source Mapping in Latin America: Collaborative Approaches in the Classroom and Field. *Journal of Latin American Geography*, 18(2), 13–32.